ESSAYS ON U.S. STATE-LEVEL FINANCIAL WEALTH DATA AND CONSUMPTION DATA

by

Xia Zhou

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

October, 2010

Abstract

A new direction in emerging research is to use geographical variations to address questions in macroeconomics that have not been resolved using aggregate data. This dissertation consists of three essays that examine geographically disaggregated U.S. data that are key to central questions in macroeconomics. The first one describes a new panel dataset for the financial wealth of U.S. states, as constructed from anonymous proprietary account-level records of geographic wealth holdings. The essay provides evidence that the new dataset is a reliable measure of stock wealth growth at the state level. It shows that the state-specific stock wealth growth is significantly correlated with the growth of local average annual income. Nevertheless, the essay finds insignificant correlations between financial wealth and both average and median housing wealth. Additionally, there is evidence that the state-level poverty rate, sex ratio, and state income from lottery sales are all associated with state-specific financial wealth growth.

The second essay constructs a state-level measure of consumption spending, one that improves significantly on existing data sources. It compares the new dataset with four other retail sales measures, and discusses their respective advantages and disadvantages as a proxy for state consumption. The essay provides evidence that the two most popular measures of retail sales involve some mysterious prior assumptions, which could contaminate studies of consumption behaviors. It also shows that the new dataset constitutes a superior measure of consumption growth. The most persuasive evidence comes from those states in which taxable retail sales are measured directly by the government.

The third essay combines the two datasets to provide new estimates of the effects on consumption from changes in the stock and housing wealth. The paper finds large but sluggish housing wealth effects. The estimated impact on consumption of a one dollar change in the housing wealth that took place two years prior is about 6 cents. However, the data show no evidence of significant stock wealth effects. Additionally, we find mixed evidence

for statistically significant differences between housing and stock wealth effects.

Advisors:

Professor Christopher Carroll

Professor John Faust

Acknowledgement

I owe my deepest gratitude to my advisor, Professor Christopher Carroll. This dissertation would not be possible without his great guidance and support. I would also like to thank Prof. Jon Faust, Prof. Laurence Ball, Prof. Louis Maccini, Dr. Michael Palumbo, Dr. Kevin Moore, and seminar participants at the Macro Lunch Seminar of the Economics Department of Johns Hopkins University for their insightful comments and discussions on earlier drafts of the essays in this dissertation. I am also grateful to Dr. Larry Taff and my other colleagues at IXI. Without their help, the construction of the financial wealth growth data would not have been possible. Finally, I would like to thank my family, especially my husband Yiduo Zhang, for their continued support and understanding, which made this dissertation feasible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Along with the skyrocketing stock market, the U.S. personal saving rate dropped from about 8 percent to near zero during the second half of the 1990s. This phenomenon sparked renewed policy and research interests in wealth effects on consumption. In 2007, we witnessed another bubble burst, this time in the real estate market. This arguably led to another collapse of the stock market, one that drove the economy into a recession. It is therefore interesting for macro-economists to investigate wealth effects by separating housing wealth from financial wealth.

There are many reasons why a financial wealth effect could be different from a housing wealth effect. For example, equity prices are much more volatile than house prices. Consequently, people may believe that financial wealth fluctuations are temporary, which according to the permanent income hypothesis, implies a lower wealth effect. On the other hand, some researchers argue that housing is essentially a consumption item. Therefore, house price increases could lead to less spending on other consumption goods for potential house buyers. Additionally, financial wealth effect could be larger due to easier access to capital gains on equities. Thus, on average, we would expect a higher housing wealth effect. There are many other arguments for or against a higher housing wealth effect, which calls for empirical studies on differential financial and housing wealth effects.

Most studies on wealth effects employ aggregate or household-level data, probably because other options are very limited. However, it is recognized that endogeneity and aggregation problems are associated with aggregate data. Some researchers also argue that the failure to find any significant housing wealth effect, or any differences between housing and financial wealth, might be because of the multicollinearity between the two at the national level. Household-level data, however, are always criticized as containing serious measurement errors.

This dissertation proposes an alternative approach – state-level data. It is believed that such data would suffer less endogeneity and fewer aggregation problems if there is enough regional variation. Mian and Sufi (2009), for example, use zip-code level data to study the consequences of mortgage credit expansions in the U.S. Within-county cross-zip code variations enable the authors to isolate the effects of credit expansions from other potential factors, such as income shocks. Meanwhile, it is arguable that state-level data constitutes a better measure than individual household-level data. The most difficult task, however, is probably to find reliable financial wealth data and consumption data at disaggregate levels. This dissertation contributes to the current literature by introducing two new measures of U.S. state-level financial wealth growth and consumption growth, respectively.

Regarding financial wealth, there are only two publicly available measures in the U.S. – the Survey of Consumer Finances (SCF) and the Flow of Funds Accounts (FFA). Both measures have advantages and disadvantages. The SCF documents very detailed financial information as well as the demographic characteristics of American households. Nevertheless, it is only carried out every three years, which makes it unsuitable for time series studies. Additionally, its relative small sample size and suppressed geographic information makes it unsuitable for studies at any disaggregate level. Last but not least, it provides no measure of consumption, thus making it almost impossible to conduct a wealth effect study. The FFA, on the other hand, is available on a quarterly basis, and is the most widely used aggregate measure of U.S. household financial wealth. However, it shares a similar limitation as the SCF – it is only available at the aggregate level.

Case, Quigley, and Shiller (2005) did pioneering work using U.S. state-level data to estimate and compare housing wealth effects and stock wealth effects. Notably, Case, Quigley, and Shiller (2005) created a measure of state-level stock wealth for the U.S. The data is, however, limited because it is necessary to make strong assumptions in order to produce state-level estimates. For instance, they have to assume a constant proportion of mutual funds out of total financial wealth over time, which in this dissertation is found to be false. In addition, for years when their state specific data is not available, they assume the same asset distribution across states. This methodology, however, implies no regional variation for most of the years of their study period. The second chapter of this dissertation describes a new panel dataset of stock wealth for U.S. states, one constructed from anonymous proprietary account-level records of geographic wealth holdings. Evidence is presented for the accuracy of the new dataset as a measure of stock wealth growth as compared to the SCF and FFA. We then document the association between financial wealth and income, housing wealth, and other demographic variables.

Regarding consumption data, other than the aggregate Personal Consumption Expenditure (PCE) data reported by the Bureau of Economic Analysis, two widely used datasets are actually available at the household-level. Nevertheless, both are subject to serious measurement errors. The Panel Study of Income Dynamics (PSID) only measures food consumption, and the Consumer Expenditure Survey (CEX), while having detailed data on household expenditures, is noisy and provides poor financial information. There are many studies that have created state-level measures of consumption growth, but they have had to resort to retail sales data. To date, five different retail sales datasets exist in the economics literature. The third chapter of this dissertation presents an updated and improved version of the data derived from one of these sources. The new dataset is compared with other existing datasets. Specifically, the related chapter focuses on one of the most commonly used data sources, investigates its methodology, and discusses its limitations compared to the new dataset when adopted for wealth effect studies.

The current literature on wealth effects is mixed. Using household-level data, Levin

(1998) found no housing wealth effect on consumption. Bostic, Gabriel, and Painter (2005) presented similar results, but find evidence for a significant and larger housing wealth effect than stock wealth effect, though only among home-owners. On the other hand, using aggregate-level data, Davis and Palumbo (2001) provided evidence for a consistent, important non-stock market wealth effect. Additionally, Carroll, Otsuka, and Slacalek (2006) report a larger immediate housing wealth effect than stock wealth effect, although the difference is found to be statistically insignificant. Case, Quigley, and Shiller (2005) are the first to investigate wealth effects using disaggregate-level data for the U.S. They find a significant housing wealth elasticity of around 5 percent, but negligible stock wealth elasticity. The fourth chapter of this dissertation contributes to the current literature by combining the two new datasets as described above so as to investigate the effects on consumption from changes in stock and housing wealth. We find a sluggish yet important housing wealth effect, but no evidence for a significant stock wealth effect. Nevertheless, there is mixed evidence for a significantly larger housing wealth effect.

Chapter 2

Measuring U.S. State Financial Wealth Growth

2.1 Introduction

In the U.S., there are only two publicly available sets of stock wealth data that are commonly used by researchers: the Survey of Consumer Finances (SCF) and the Flow of Funds Accounts (FFA). The SCF is a triennial survey of detailed financial information, inclusive of demographic characteristics about individual U.S. households. Many researchers have documented patterns of financial wealth accumulation using the SCF. However, the SCF is only carried out every three years, making a time series study of financial wealth growth based on the SCF impossible. Furthermore, the survey interviews roughly only 4000 households and withholds all geographic information about the interviewees. This makes any study at an aggregate geographic level other than the national one impossible. The FFA, on the other hand, consists of the most widely used aggregate data available on U.S. household balance sheets provided on a quarterly basis. It has been used by macro-economists for time series studies on wealth accumulation. However, as with the SCF, researchers using the FFA data are restricted to national level studies only. Consequently, any dataset that expands on the limitations of the SCF and the FFA would greatly contribute to studies related to stock wealth, and would thus be welcomed by researchers.

Case, Quigley, and Shiller (2005) appears to be the only paper that utilizes data sources other than the SCF and FFA. They created a measure of state-level stock wealth for the U.S. The data is limited, however, because strong assumptions have to be made in order to produce state-level estimates. For instance, the authors assume a constant share of mutual funds out of total financial assets.[1]

This paper describes a new panel dataset of stock wealth for U.S. states, one constructed from anonymous proprietary account-level records of geographic wealth holdings. It further provides evidence that the new dataset constitutes a very good alternative measure of financial wealth growth as compared to the SCF and FFA. The rest of the paper is organized as follows: Section 2 describes the data source for the new dataset, and how the dataset is constructed. Section 3 examines the validity of the new dataset by comparing it with the SCF and FFA at the aggregate level. Section 4 explores the existence of geographical patterns underlying stock wealth growth across states, and investigates the relationship between specialization and stock wealth growth. Section 5 studies the relationship between stock wealth, income and housing wealth using the new dataset. Section 6 explores the association between stock wealth and selected demographic variables. Section 7 concludes.

2.2 Data description

The data described in this chapter is a measure of the stock wealth of U.S. states, with a semiannual frequency over the period 2000 – 2005. To be consistent with the FFA, stock market wealth is defined as the sum of directly and indirectly held (those invested in IRA and Keogh accounts) stocks and mutual funds. The data were constructed while the author was employed part-time at the IXI Corporation over a period of two years. Conditional on that employment, IXI permitted the author to produce the stock wealth data at the U.S. state level.

[1] Please refer to Chapter 4 of this dissertation for a detailed discussion concerning the limitations of their data.

The state-level data was generated using semiannual anonymous account-level records of financial wealth holdings at the ZIP+4 code level from the IXI Corporation, through its network known as IXI▶Net^{TM}. The IXI▶Net^{TM} contains data on more than 85 leading financial institutions, inclusive of major banks, brokerage firms, insurance companies, and mutual funds dealers. The actual names of the financial institutions reporting to IXI are suppressed and cannot be revealed. IXI receives absolutely no non-public, personally identifiable information concerning U.S. households. Additional information can be found on their web site: www.ixicorp.com.

At the end of each semiannual cycle, IXI collects data from financial institutions in its network, IXI▶Net^{TM}. In order to better assure data confidentiality, once the ZIP+4 code level data is received, it is joined and averaged with other ZIP+4 codes. The average value is then imputed back down to the ZIP+4 codes participating in the respective averaging. The rules utilized in performing this joining and averaging are complex, and to some extent are based on geographical proximity. The most important factor is the requirement that there should be at least 7 households for each ZIP+4 code. Thus, a ZIP+4 code with fewer than 7 households will be joined with its neighbors in order to achieve the requirement.

The number of institutions in the IXI network is constantly changing over time. Hence, the biggest challenge in constructing the stock wealth dataset was the possibility that variations in wealth would be caused by factors other than wealth holding variations of state residents. To illustrate, assume company A and company B report to IXI at time $t = 0$, companies A and C report at time $t = 1$, and companies A, B, and D at time $t = 2$. Also, assume for the time being that companies A, B, C, and D are four different companies. When calculating the growth rate of financial wealth at $t = 1$, all data reported by company B and company C have to be dropped. Otherwise, the calculated growth rate of financial wealth would partly reflect the change of IXI network. Similarly, the growth rate of financial wealth for $t = 2$ is calculated using data from company A and only company A again. Company mergers and splits, however, make this problem much more complicated. For example, if company A acquires company B at time $t = 1$, the above example would now require data from company B for the calculation of the financial wealth growth rate at time $t = 1$. On the other hand, if company A splits into company A and company D at time $t = 2$, data from company D will be combined with data from company A at $t = 2$ when compared with data from company A alone at time $t = 1$. There are many other reasons that could further complicate the problem. To minimize the problem, a great deal of effort was expended in tracking all mergers and institutions' membership in the network.

The formation of a consistent source of financial data over time was carried out at the state level. Thus, a common group of reporting institutions was constructed for every TWO ADJACENT HALF YEARS. The growth rates could therefore be calculated on a rolling basis as the log difference of two adjacent values. In other words, total assets from institutions reporting to IXI at both time t and $t-1$ for state i were first summed up; the log difference of the sums were then taken as the growth rate of the stock wealth of state i at time t.

Specifically, suppose that $F^j_{i,t}$ is the total amount of stock wealth reported by institution j at time t for state i. $\Omega_{i,t}$ is the set of all institutions reporting at time t for state i. $\Omega_{i,t} \bigcap \Omega_{i,t-1}$ is then the set of all institutions reporting for state i at both time t and $t-1$. $\Delta F_{i,t}$, the growth rate of stock wealth of state i at time t is defined as:

$$\Delta F_{i,t} = \ln\Big(\sum_{j \in \Omega_{i,t} \cap \Omega_{i,t-1}} F^j_{i,t}\Big) - \ln\Big(\sum_{j \in \Omega_{i,t} \cap \Omega_{i,t-1}} F^j_{i,t-1}\Big), \tag{2.1}$$

where $i = 1, 2, ..., 51; t = 2000h2, 2001h1, ..., 2005h2$.

After obtaining the correct growth rates, the total stock market wealth for each state is imputed backwards as:

$$\hat{F}_{i,T} = F_{i,T} \tag{2.2}$$

$$\hat{F}_{i,T-1} = \hat{F}_{i,T} / \exp(\Delta F_{i,T}) \tag{2.3}$$

$$\hat{F}_{i,T-2} = \hat{F}_{i,T-1} / \exp(\Delta F_{i,T-1}), \tag{2.4}$$

....

where $F_{i,T} = \sum_{j \in \Omega_{i,T}} F^j_{i,T}$, and $T = 2005h2$.

As will be shown later in this chapter, IXI only measures a proportion of the total stock wealth of American households as measured by the SCF and FFA. Thus, it is assumed for all states and time periods that $\frac{F_{i,t}}{F^*_{i,t}} = \frac{F_{\text{IXI},t0}}{F_{\text{FFA},t0}}$, where $F_{\text{IXI},t0} = \sum_i F_{i,t0}$ and $F_{FFA,t0}$ is the total stock wealth as measured by the FFA at $t0 = 2000h2$ – the real state stock wealth at time t can be derived as $F^*_{i,t} = F_{i,t} * \frac{F_{\text{IXI},t0}}{F_{\text{FFA},t0}}$, where $F_{\text{IXI},t0} = \sum_i F_{i,t0}$.

Real stock wealth per capita is then defined as:

$$\hat{f}_{i,t} = \hat{F}_{i,t}/(\text{CPI}_t/\text{POP}_{i,t}), \tag{2.5}$$

where CPI_t is the consumer price index at time t, and $POP_{i,t}$ is the population of state i at time t. The growth rate of real stock wealth per capita is thus calculated as:

$$\Delta f_{i,t} = \ln(f_{i,t}) - \ln(f_{i,t-1}). \tag{2.6}$$

The new data, as described above, might be different from the growth rate of the true financial wealth. For example, it is possible that the data might miss some of the very wealthiest households if their assets are likely to be held at “boutique” institutions or other sources from which IXI does not collect data. Additionally, the market share of a financial institution changes over time. This variation will inevitably be reflected in our measure of financial wealth growth. The author, however, does not have any way of measuring how large these problems might be as there are no other alternative data sources at the U.S. state level to compare with. The following section, therefore, provides evidence for the validity of the new data at the aggregate level.

2.3 The new data versus the FFA and SCF

This section examines the validity of the new measure of financial wealth growth at the aggregate level. Specifically, we prove the credibility of the new dataset by comparing it with the FFA and SCF, which are jointly considered as the "gold-standard" or "benchmark" in the current literature. This section first discusses the FFA and SCF, and their corresponding advantages and disadvantages as data sources. The two are then compared with the new dataset.

2.3.1 The FFA and SCF: why they are different.

Researchers commonly use FFA data as an aggregate measure of wealth. Concerns remain, however, regarding its accuracy as a measure. For instance, the financial assets and liabilities of the household sector as measured by the FFA are derived as residuals, because the related activities on the balance sheets of American households are unavailable. More specifically, it is derived by deducting the activities listed under other sectors on the balance sheets from nationwide totals. The concern then is that such a method could potentially introduce large errors where the residual part is small as compared to other sectors.[2]

The SCF, on the other hand, is the most comprehensive survey available on the individual assets and liabilities of American households. In light of the highly skewed distribution of wealth across American households, a two-step sampling method – a random sampling and an over-sampling of rich people[3] – is adopted in order to avoid the large bias found in other surveys, such as the Survey of Income and Participation Program (SIPP), the Panel Study of Income Dynamics, and the Consumer Expenditure Survey. A nonresponse-adjusted sampling weight is assigned to each interviewee according to his or her income and demographic characteristics. An aggregate estimate of total wealth is calculated as the weighted sum of all responses.[4] Despite the careful survey design and rigorous weighting scheme, the SCF is not immune to various issues, among them, misunderstandings regarding the questionnaire,

[2]Curtin, Juster, and Morgan (1989) provides further discussion on this issue.
[3]See Kennickell (2006c) for further discussion.
[4]See Kennickell (1999).

data imputation due to non-responses or problems related to identity confidentiality, and the use of inaccurate information in the over-sampling of wealthy households.[5]

Apart from the different data sources being utilized, several other major factors result in differences between the FFA and SCF. First, since 2001, the FFA has not separated nonprofit organizations from the household sector. According to the data reported in 2000, nonprofit organizations accounted for roughly 6 percent of total financial assets and 7 percent of total stock wealth in both sectors. Figure 2.1 plots the share of stock wealth of nonprofit organizations over time. It is obvious that although the wealth of nonprofit organizations constitutes only a small component, it fluctuates over time as a percent of the total wealth in household and nonprofit organization sectors. It is therefore difficult to accurately single out stock wealth in household sectors only. Second, assets in IRA and KEOGH accounts from the FFA are listed in the corresponding asset categories where the money is actually invested. For example, all IRA and KEOGH accounts invested in mutual funds are listed in the "mutual funds" category and cannot be distinguished from other types of mutual funds indicated in the FFA. In the SCF, however, IRA and KEOGH accounts are treated as constituting one asset category, one where limited information regarding how a specific investment is invested is available. Before 2004, interviewees were asked if most of their money in IRA and KEOGH accounts was invested in bank accounts, stocks and mutual funds, bonds, any combination of the above, or others.[6] In order to make measures from the FFA and SCF comparable, certain assumptions have to be made for the SCF when allocating IRA and KEOGH accounts to different asset categories. Throughout this paper, equal proportions were assumed whenever the response was a combination of more than one type of asset.

2.3.2 Is the new data close to the benchmark?

In order to investigate the validity of the new dataset as a measure of stock wealth at the national level, we first need to study the coverage of the new dataset as compared to

[5] See Kennickell (2006b); also Moore and Johnson (2005).

[6] From 2004 onward, interviewees were only able to choose between "all in stocks," "all in interest earning assets" and "split."

the FFA and SCF. The sum of the stock wealth from all of the institutions reporting to IXI at the end of the second half of 2001 was calculated and compared to the aggregate stock wealth as measured by the FFA and SCF for the same time period. The year 2001 was chosen for two reasons. First, the FFA discontinued reporting a separate measure for household assets values at the end of 2000. In order to project the stock wealth measured by the FFA for household sectors after 2000, this paper must assume that the proportion of stock wealth held by nonprofit organizations remains at the year 2000 level from that year onward. Therefore, picking a time period that is close to the year 2000 would be preferable. Second, beginning in 2004, the SCF questionnaire became less clear in terms of how it broke IRA/KEOGH accounts into specific asset categories. As a result, stocks and mutual funds that are held indirectly cannot be clearly separated in the form of retirement accounts from other asset categories; this diminishes the comparability of the SCF from 2004 onward.

Coverage of the IXI data at the national level is presented in Table 2.1. The projected FFA[7] and SCF numbers at the end of 2001 serve as benchmarks. The Table suggests that the IXI data is estimated as covering more than 30 percent of U.S. aggregate stock holdings and more than 50 percent of aggregate mutual fund holdings. Using my definition of stock market wealth, IXI is estimated to aggregate data of roughly 40 percent of the total U.S. stock market assets held by U.S. households. Table 1 provides encouraging evidence that the new dataset has great potential as a measure of stock wealth.

This paper further investigates whether the new dataset represents stock wealth growth at the aggregate level. Aggregate stock wealth growth rates, as measured by both the FFA and the new dataset between 2000h1 and 2005h2, are presented in Figure 2.2. Please note that for the FFA measure, total stock wealth for both the household sector and nonprofit organization sector is used under the assumption that the wealth held by nonprofit organization sectors constitutes the same proportion over time. Despite using completely different data sources, Figure 2.2 shows that the two series move similarly to one another, suggesting

[7]Thanks to Dr. Michael G. Palumbo, who kindly provided detailed information about the components of each FFA asset category.

that the new data is representative of the nation as a whole.

The discussion above provides evidence that the new dataset is representative at the national level. See the fourth chapter of this dissertation regarding the validity tests of the data at the U.S. state level. After proving the validity of the new dataset, the rest of this paper then investigates the characteristics of the financial wealth growth at the U.S. state level, which could not be performed before.[8]

2.4 The regional effect?

Economists have developed a rich body of literature on geographic economics, one that mainly focuses on productivity. This section explores whether there are also large differences in wealth accumulation patterns across states, additionally whether any existing agglomeration theories might be useful in explaining any differences.

Figure 2.3 plots the distribution of the half annual growth in stock wealth across states in the U.S., from 2001h1 to 2005h2. Similar patterns are found across states, something well expected as the U.S. stock market is well integrated. To account for this, stock wealth growth at the state level, $\Delta w^f_{i,t}$, is decomposed into an aggregate component, $\Delta w^f_{US,t}$, and a state-specific component, $\Delta \tilde{w}^f_{i,t} = \Delta w^f_{i,t} - \Delta w^f_{US,t}$. As indicated in Figure 2.3, the aggregate component is the dominant determinant of state-level stock wealth growth. The state-specific component, however, is expected to vary across states, and is the component in which we are primarily interested. Figure 2.4 shows the distribution of state-specific stock wealth growth across states. It shows that the state specific wealth growth, $\Delta \tilde{w}^f_{i,t}$, does differ across states, much more than $\Delta w^f_{i,t}$ does. Figure 2.4 also shows significant half annual seasonal patterns, and provides supporting evidence for the need to use annual growth rates, as discussed in the fourth chapter of this dissertation, where we investigate wealth effects.

[8] The data archive that can produce all results in this study is available from Johns Hopkins library, at URL: http://jhir.library.jhu.edu/handle/1774.2/34267. Instructions on how to obtain the new data of financial wealth growth rate for U.S. states can be found in the read me file for the data archive.

An interesting question is whether $\Delta\tilde{w}^f_{i,t}$ exhibits geographical patterns; i.e., do neighboring states share similar stock wealth accumulation patterns? To answer this question, we group all 50 states plus Washington, D.C. into nine census divisions. To better understand whether states within the same census division share similar wealth growth patterns, we regress $\Delta\tilde{w}^f_{i,t}$ on census division dummies, with and without a time dummy. Nevertheless, Table 2.2 shows that for both regressions, regional dummies are not significant, suggesting that there is no statistically significant average regional effect.

It remains an interesting question then as to what drives differences in stock wealth growth across states. Three main sets of theories have been proposed to explain the clustering effect observed in productivity. Although conflicting with one another, all three deal with the externality associated with specialization.[9] It is widely accepted by researchers that insuring production risk promotes specialization. Risk sharing through capital markets, which requires well diversified portfolios, is considered to be one way of insuring against production risk. This theory, however, suggests that specialization is linked to variance rather than the level of stock wealth growth.

Following the literature, state specialization in a specific industry is defined as the fraction of that industry's share of that respective state's employment, relative to its national share of U.S. employment. Specifically, the specification index for state i in industry j is $SPEC_{i,j} = \frac{E_{i,j}/E_i}{E_{\text{U.S.},j}/E_{\text{U.S.}}}$, where $E_{i,j}$ denotes the level of employment in state i for industry j, E_i total employment in state i, and $E_{\text{U.S.},j}$ and $E_{\text{U.S.}}$ their counterparts at the national level. Therefore, state i's specialization is measured as the state's average specialization for all major industries, i.e., $SPEC_i = \sum_{j=1}^{n} SPEC_{i,j}/n$. Employment data at the 2-digit manufacturing industry level is obtained from the 2002 Economic Census, and is used to calculate the specialization index. When an employment number for a specific industry and state is not disclosed for confidentiality reasons, the midpoint of the provided range is used in the calculation. For example, if employment is suppressed, but is reported to fall

[9] See Wagner (1891), Arrow (1962), Romer (1986), Porter (1990), and Jacobs (1969)

between 500 and 999, 750 is used as the real employment number.

Figure 2.5(a) plots the mean specialization index for each state for 2002 versus the volatility of $\Delta\tilde{w}^f_{i,t}$ for each state between 2001 and 2005 – $\sigma(\Delta\tilde{w}_{i,\cdot})$. The solid line is the fitted value of $\sigma(\Delta\tilde{w}_{i,\cdot})$, derived by regressing $\sigma(\Delta\tilde{w}_{i,\cdot})$ on $SPEC_i$ and a constant term only. The plot shows a positive yet weak correlation between $\sigma(\Delta\tilde{w}_{i,\cdot})$ and $SPEC_i$. Washington, D.C. stands out in Figure 2.5(a) as an outlier. It is unique in the sense that the government and government enterprise sectors, as opposed to manufacturing industries, constitute the largest proportion of its employment and production. Figure 2.5(b) consequently raises a valid concern that the problematic specialization index for Washington, D.C. could sabotage our investigation of the correlation between stock wealth growth and specialization. However, even after Washington, D.C. is removed from the sample, Table 2.3 confirms that the level of specialization does not help explain the variance in stock wealth growth. Additionally, Table 2.3 suggests that income growth, whether measured by level or variance, cannot help us predict stock wealth beyond the level of specialization index.

2.5 Income and housing wealth

2.5.1 Average income and median income

Many studies have investigated the relationship between income and wealth using the SCF. However, the SCF does not survey the same set of consumers over time, and is only carried out every three years, thus making a time series study at disaggregate levels difficult. This paper further explores the relationship between income and wealth using the new U.S. state-level data. The goal is to find evidence either in support of or against the current literature, something that cannot be done with other datasets.

State average income is obtained from the Bureau of Economic Analysis (BEA); real average income $y^a_{i,t}$ is then calculated by dividing state income by respective state populations and the national CPI. Note that $y^a_{i,t}$ is not a comprehensive measure of income. It only includes earnings based on participation in production – wages, salaries, rental income and so forth – while excluding net gains from the sale of assets, pension benefit payments, and personal contributions to social insurance.[10] $y^m_{i,t}$ measures the state pre-tax median household income as reported by the Census Bureau's American Community Survey.

Figure 2.6 examines the correlation between the annual growth of stock wealth Δw^f_i, and the annual growth of average income $\Delta y^a_{i,t}$ and median income $\Delta y^m_{i,t}$.[11] Indexing states by i, the table at the bottom of Figure 2.6 suggests that the growth of average income significantly and positively correlates with stock wealth growth. It further shows a positive but weak and insignificant correlation between median income growth and stock wealth growth.

This observation is consistent with what current studies have documented regarding wealth and income inequalities. First, net worth is much more unequally distributed than income. Kennickell (2006a) shows that from 1989 to 2004, the wealth Gini coefficient rose steadily from 0.7863 in 1989 to a statistically significant 0.8047 in 2004. Over the same period, however, the coefficient for income fluctuated around 0.53, and ended up in 2004 at a similar level as in 1989. Zhu (2007) summarizes the average net worth and income of American households in the SCF for 1995 and 2004 by quartile, and provides supporting evidence for greater wealth inequality than income disparity. Second, about 56 percent of total income in the U.S. is earned by those in the top income quartile. Therefore, a large portion of state average income growth is driven by the top earners in the state, those individuals more likely to be stock owners and to invest income gains in various types of assets. According to the SCF for 2004, about 57 percent of households in the first income quartile are stock owners; collectively, they hold more than 80 percent of total U.S. stock wealth. Therefore, if average income growth mainly reflects the income growth of those in the top tier of income distribution, it is reasonable to expect a positive correlation between income growth and stock wealth growth. Using the SCF for 2004, Table 2.4 shows much

[10]The growth of average labor income was also tested, but the results were suppressed because of similarity.

[11]To be consistent with the fourth chapter of this dissertation, the annual stock wealth growth is constructed by taking the log differences between the stock wealth at the end of the first half of each year.

lower stock ownership among households in the lower half of income distribution. Median income growth then may not contribute significantly to stock wealth growth. This means that, even when median income growth reflects income redistribution or an improvement in income inequality, one might still observe stock wealth growth moving in the opposite direction of median income growth.

2.5.2 Housing wealth

For most American households, their homes constitute a greater chunk of their total wealth than stock wealth does. There are many theoretical reasons as to why changes in housing wealth should have a different impact on consumption than stock wealth. Some empirical studies, using both aggregate and household level datasets, have investigated the differential effects of housing and stock wealth on consumption. The fourth chapter of this dissertation contributes to the existing literature by using state-level data to provide evidence that housing wealth has a greater effect than stock wealth on consumption. This paper then uses state-level data to explore the relationship between the two respective growth rates.

According to the SCF for 2004, 69 percent of American households are homeowners; 37 percent of these are also stock owners, who collectively hold roughly 97 percent of the total stock wealth in the U.S. There is, however, survey-based evidence for weak links between disaggregate housing wealth growth and stock wealth growth. The SCF 2004 found that 34 percent of those homeowners who had refinanced their first-lien mortgage during the previous three years borrowed more than the amount being refinanced. Nonetheless, home improvements and debt relief accounted for 45 and 31 percent, respectively, of the amount of equity extracted from residential real estates. A survey conducted by the National Association of Realtors also found that for repeat buyers, most of the capital gains from selling a home were spent on down payments or home improvements for new homes. In addition, a majority of funds spent on unrealized gains were used for debt relief. Only 5 percent of such funds were saved in the bank. Nevertheless, Green (2002), by using county level data in California, found that changes in stock price could "cause" house prices to change.

This conclusion arguably suggests that stock wealth and house wealth are highly correlated.

Two different measures of housing wealth are used in this paper: average housing wealth, $w_{i,t}^{h,a}$, and median housing wealth, $w_{i,t}^{h,m}$. The formula used for constructing state-level average housing wealth is similar to the one utilized by Case, Quigley, and Shiller (2005), and is given as follows:

$$w_{i,t}^{h,a} = (HO_{i,t} * HH_{i,t}) * HPI_{i,t} * HV_i / POP_{i,t},$$

where HO is the home ownership rate taken from the Census Bureau; HPI, the weighted repeat sales housing price index from the Federal Housing Finance Agency (FHFA); HV, the average home price for 1999 from the 2000 Census; and $POP_{i,t}$, the state population.[12] Median housing wealth is reported by the American Community Survey (ACS).

Figure 2.7 visually shows a weak correlation between stock wealth growth and housing wealth growth, whether based on the average or the median. The table at the bottom of Figure 2.7 further documents the insignificance of the correlation. The lack of a correlation between stock wealth growth and housing wealth growth proves the validity of separating the respective effects as measured in the fourth chapter. This observation is consistent with the above noted survey-based evidence that American households only financially invest a small amount of their unrealized capital gains. It is, however, in contrast with what Green (2002) found.[13]

Nevertheless, the author recognizes that this positive yet insignificant correlation coefficient could be caused by our short sample period, 2001 to 2005. This specific time period is abnormal in the sense that the U.S. stock market plunged in 2001 and continued to slide over the next couple of years; the U.S. housing market overheated during the same period, leading to a serious housing market crash in 2007. Taking that into consideration, the au-

[12] Please refer to Chapter 4 of this dissertation concerning the data.

[13] So as to be consistent with the method employed in Green (2002), this paper also investigates the correlation between lagged stock wealth growth and current housing wealth growth. Similar results were found, and thus are not reported here.

thor would be very much interested in repeating the same exercise when more data becomes available.

2.6 Is the growth of stock wealth linked to other demographic variables?

Since the SCF includes rich information on demographic characteristics, many researchers have used it to document wealth distributions across various demographic groups for different years. However, to the best of my knowledge, no study has used any disaggregate dataset to investigate the time series behavior of those variables related to stock wealth growth. This section studies the association of several of the most discussed variables with respect to stock wealth growth.

The selected variables include: %Poverty – the percentage of people over 16 years old under the poverty line; %Bachelor Degree – the percentage of the 25+ population who have completed a bachelor degree; %Home Language not English – the percentage of people 5 years old and over who speak a language at home other than English; Sex-Ratio – the male to female ratio; Age-Dependency – the percentage of the combined child (under 14 years old) and aged population (65 years old and over); %White – the percentage of non-Hispanic whites; %Black – the percentage of non-Hispanic blacks; Lottery – state-administered lottery ticket sales per capita (in thousands);[14] and Health Expenditure – total personal health care (PHC) expenditures per capita (in thousands.)

All the above mentioned variables, apart from Lottery and Health, are measured by the American Community Survey (ACS) of the Census Bureau. State-level lottery data is reported by the Census Bureau in the State Government Finance section. Estimates of PHC expenditures are derived from the National Health Expenditure Accounts and are reported by the Centers of Medicare and Medicare Services.[15] Furthermore, all variables derived

[14] Excluding the commissions made from selling lottery tickets.
[15] Available at http://www.cms.hhs.gov/NationalHealthExpendData/.

from the ACS are available for all 50 states plus Washington, D.C. for 2001 to 2005. The most recent PHC data at the state level is for the year 2004; lottery data is neither available for all states, nor for all consecutive periods for those states where it is available.[16] Table 2.6 documents the summary statistics for all of the variables.

Column 1 in Table 2.5 examines the correlation between the state-specific growth rates of stock wealth and the above mentioned variables. Columns 2 and 3 present the correlation between the state average demographic level across years, and the average and standard error of stock wealth growth respectively. Five variables are found to correlate in some form with financial wealth growth. These are age-dependency, sex-ratio, blacks, lottery, and health expenditure.

As expected, poverty, whether measured by level or growth, is negatively associated with stock wealth growth. Furthermore, higher poverty levels can be linked to greater fluctuations in stock wealth growth.

States with relatively greater stock wealth growth are found to have higher education levels. Conversely, states that have more people that speak a second language accumulate wealth more slowly than others. One probable explanation could be that the greater number of second language speakers is due more to immigration than education, and the fact that immigrants are more likely to be risk averse, thus implying lower risk premiums.

Surprisingly, an increase in the male to female ratio is found to negatively correlate with state-specific stock wealth growth. Furthermore, a higher male to female ratio is associated with less fluctuations in financial wealth growth.

The paper shows that a higher percentage of child and aged population is associated with lower state-specific stock wealth growth. A probable explanation for this observation

[16]Lottery income data do not exist for Alabama, Alaska, Arkansas, Hawaii, Mississippi, Nevada, North Carolina, Oklahoma, Utah, or Wyoming. North Dakota, South Carolina, and Tennessee do not have data for the whole period investigated in this paper.

would be that a household with a higher age-dependency rate is more conservative in terms of financial investment. Another possible reason would be that less money can be set aside each year for financial investment in a state where a smaller proportion of the population is made up of wage earners.

This paper also found a significant correlation between greater fluctuations in financial wealth growth and higher lottery ticket sales. This interesting observation might come from two directions. First, more risk-seeking individuals are more likely to have higher stock ownership, likewise to hold riskier stocks; at the same time, such risk seekers are more likely to gamble – that is, to buy lottery tickets. On the other hand, greater uncertainty might cause people to resort more readily to gambling.

Another interesting finding is that people who accumulate wealth more rapidly than others may not increase their spending on health-related products significantly faster than others. Nonetheless, they do have greater health expenditures as their wealth grows faster.

2.7 Conclusion

This paper describes a new dataset of stock wealth at the U.S. state level. By comparing it to the SCF and FFA, this paper provides evidence that the new dataset could be a reliable measure of stock wealth growth at the state level. This paper fails to find evident proof for geographic patterns in stock wealth growth, likewise evidence that specialization is a cause of differences across states in stock wealth variations. Average income growth, not median income growth, is found to be highly and significantly correlated with financial wealth growth. However, neither mean nor median housing wealth growth is found to correlate with state-specific financial wealth growth. This paper also presents evidence for an association between several demographic variables and the level of, or variance in, financial wealth growth. It is well recognized that the sample period of the new dataset is short, which might diminish the robustness of some of the above mentioned findings. It would be very interesting to see if the findings in this paper will remain unchanged when data for a longer period becomes available.

Figure 2.1: The share of stock wealth held by non-profit organization sectors

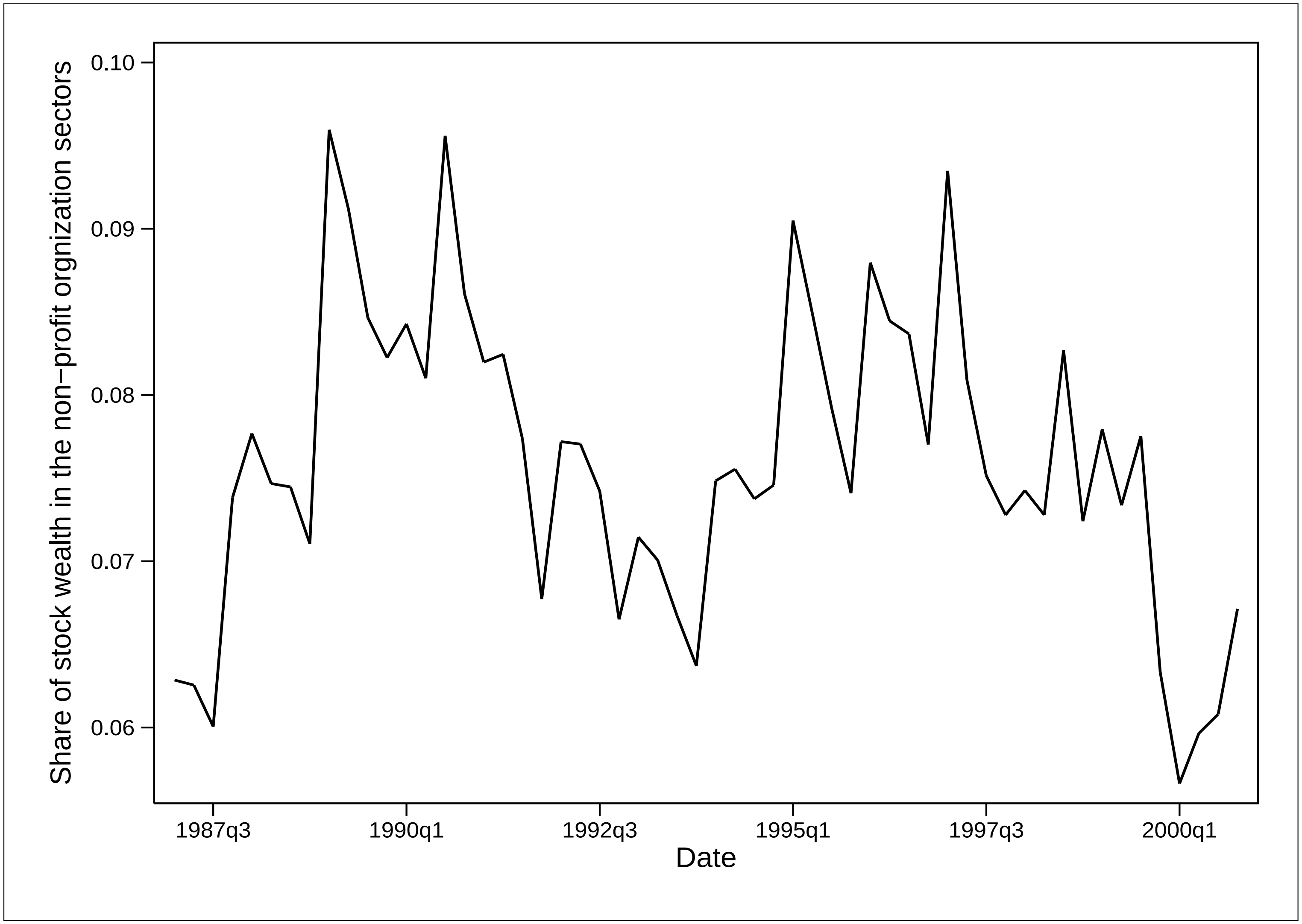


Figure 2.2: IXI versus the FFA

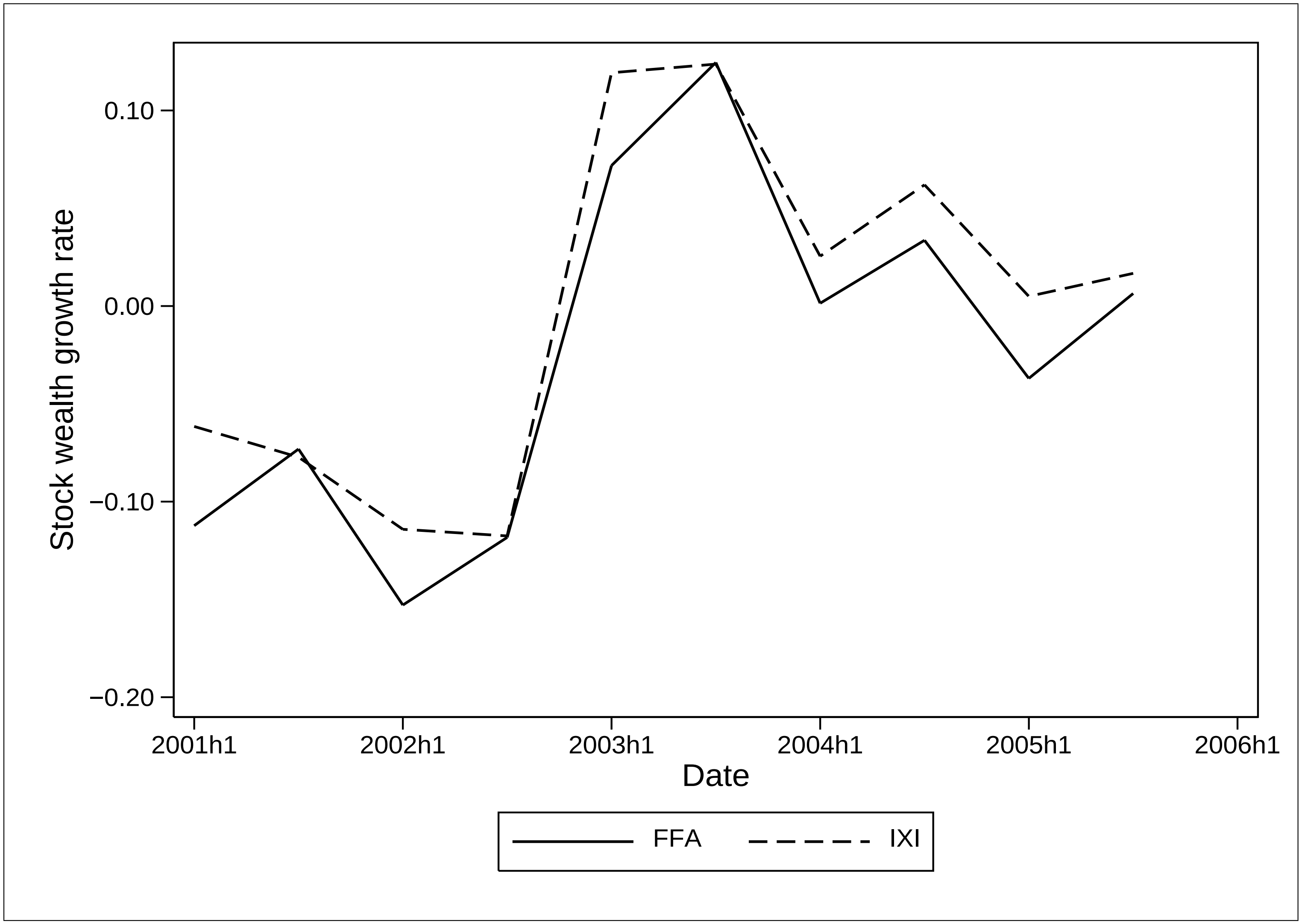


Figure 2.3: Wealth growth rates across states: $\Delta w_{i,t}^{f}$

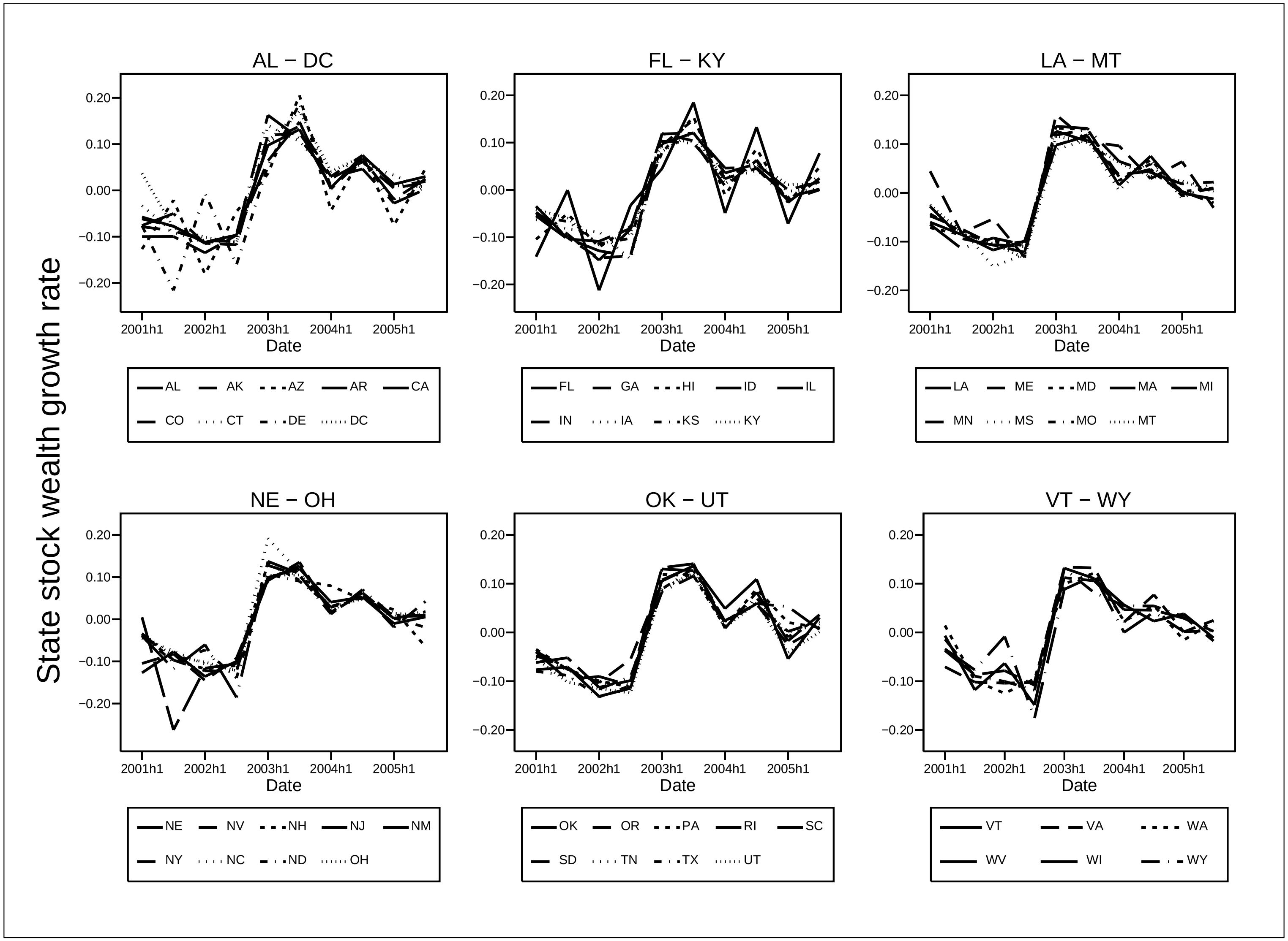


Figure 2.4: State-specific wealth growth across states: $\Delta\tilde{w}^f_{i,t} = \Delta w^f_i - \Delta w^f_{U.S.}$

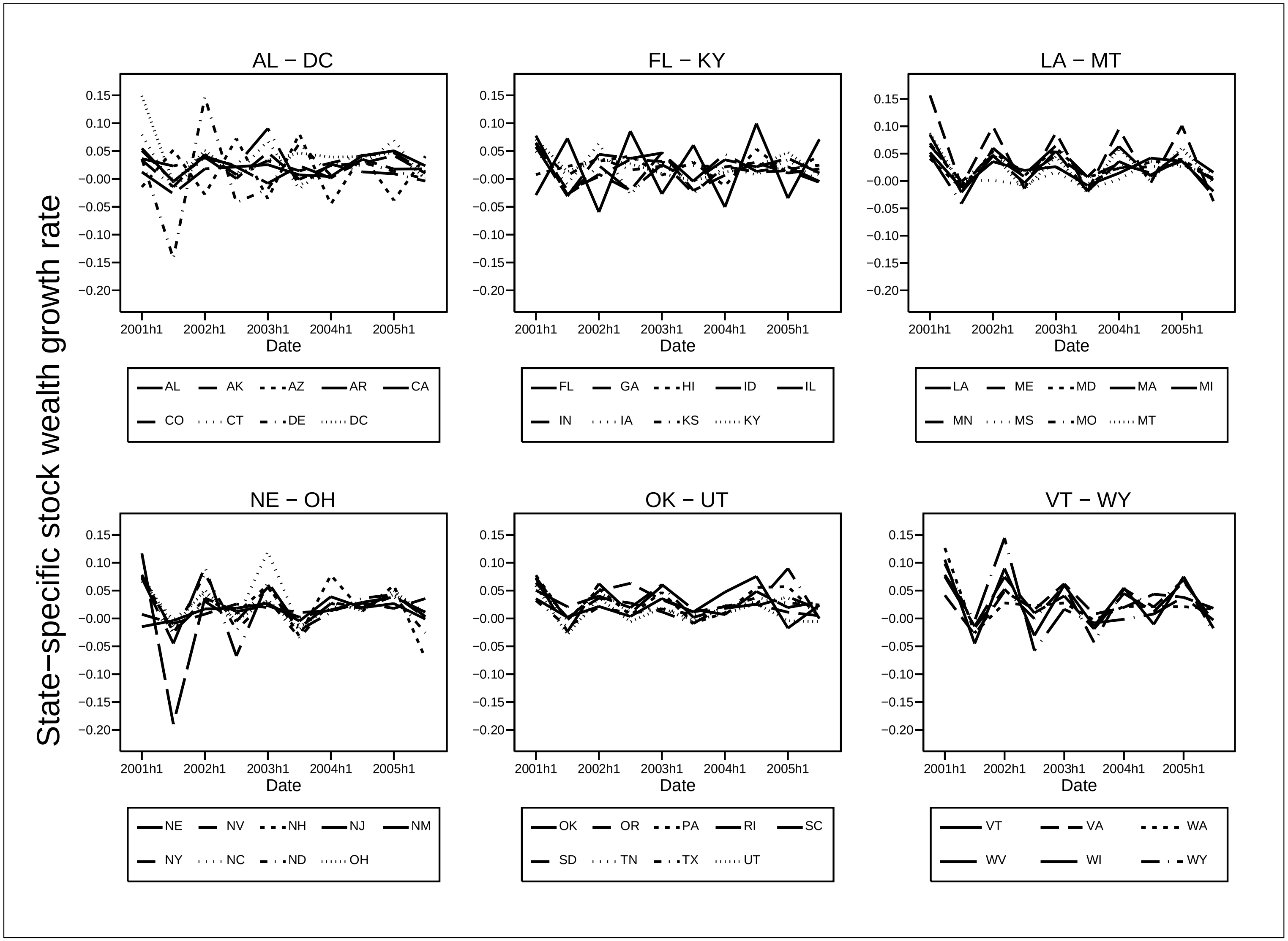
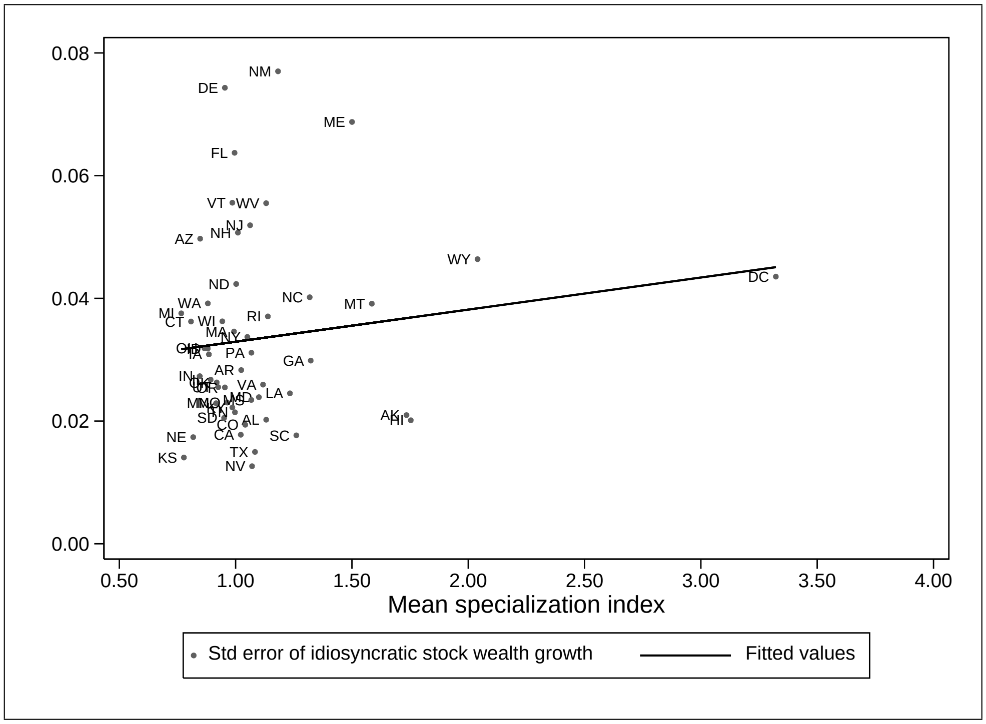
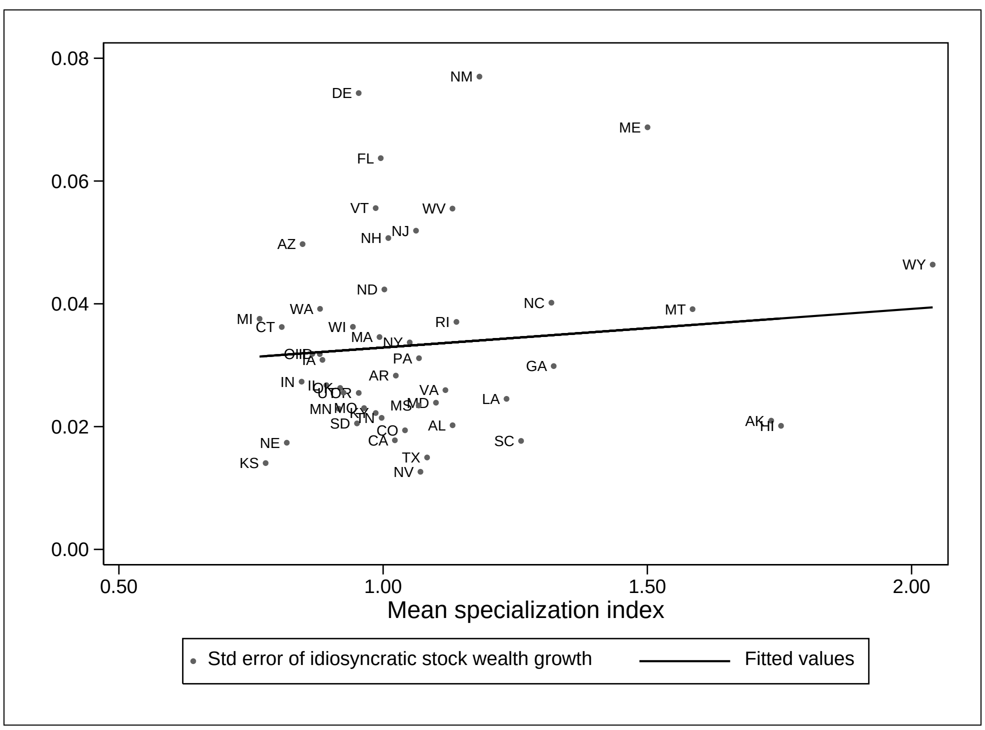


Figure 2.5: Mean specialization index versus $\sigma(\Delta \tilde{w}^{f}_{i,\bullet})$

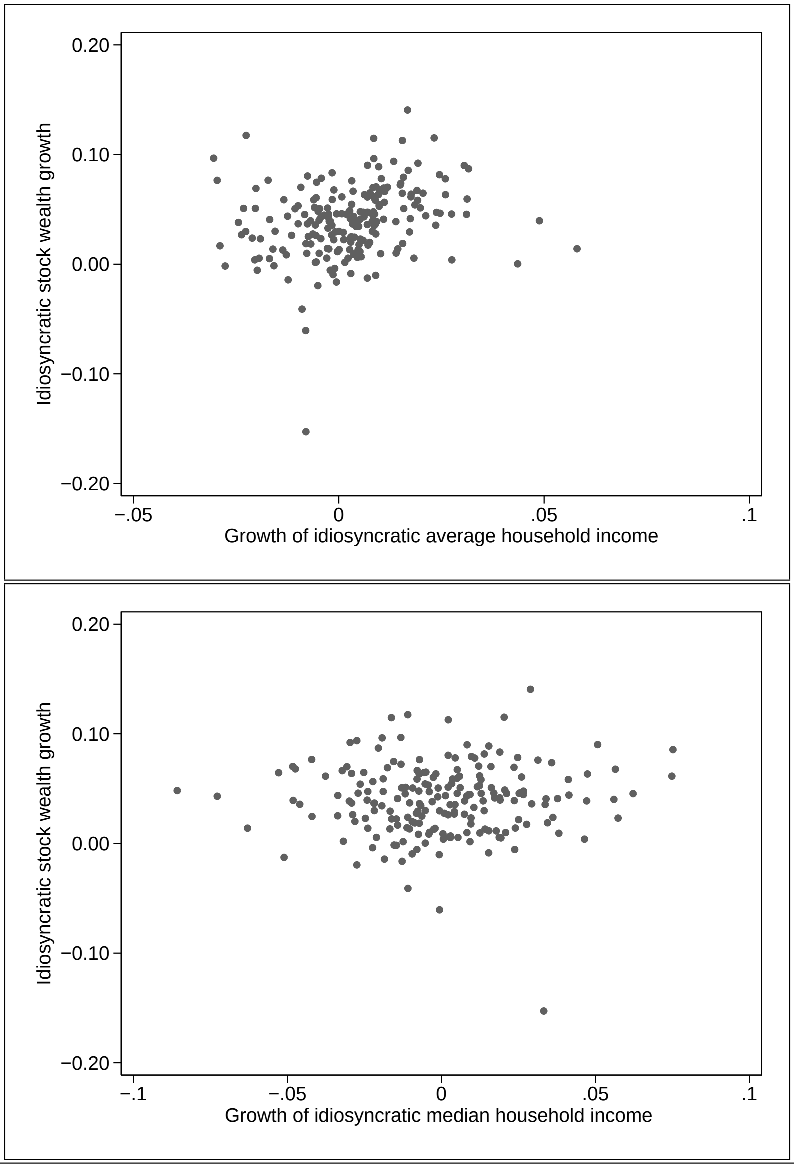


(a) All states



(b) Washington, D.C. excluded

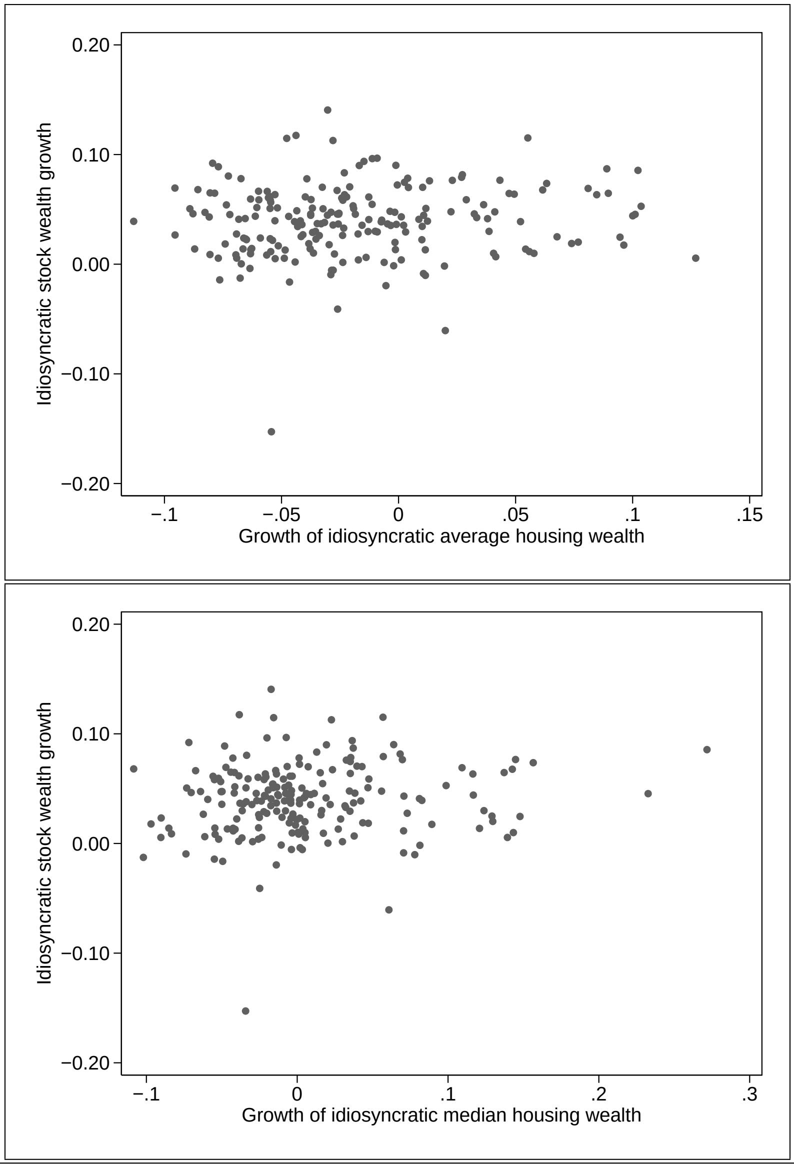
Figure 2.6: State-specific stock wealth growth versus state-specific income growth



		$\Delta\tilde{y}^a_{i,t}$	$\Delta\tilde{y}^m_{i,t}$
$\Delta\tilde{w}^f_{i,t}$	ρ	0.234^{***}	0.034
	(std error)	(0.068)	(0.070)

$t = 2001 - 2005$; number of observations: 255.
$\Delta\tilde{y}^a_{i,t}$: annual growth of state-specific average income ($\Delta y^a_{i,t} - \Delta y^a_{US,t}$).
$\Delta\tilde{y}^m_{i,t}$: annual growth of state-specific median income ($\Delta y^a_{i,t} - \Delta y^a_{US,t}$).

Figure 2.7: State-specific stock wealth growth versus state-specific housing wealth growth



		$\Delta\tilde{w}_{i,t}^{h,a}$	$\Delta\tilde{w}_{i,t}^{h,m}$
$\Delta\tilde{w}_{i,t}^{f}$	ρ	0.087	0.109
	(std error)	(0.070)	(.070)

$t = 2001 - 2005$; number of observations: 255.
$\Delta\tilde{w}_{i,t}^{h,a}$: annual growth of state-specific average housing wealth ($\Delta w_{i,t}^{h,a} - \Delta w_{US,t}^{h,a}$).
$\Delta\tilde{w}_{i,t}^{h,m}$: annual growth of state-specific median housing wealth ($\Delta w_{i,t}^{h,a} - \Delta w_{US,t}^{h,a}$).

Figure 2.8: Growth of stock wealth versus growth of demographics variables

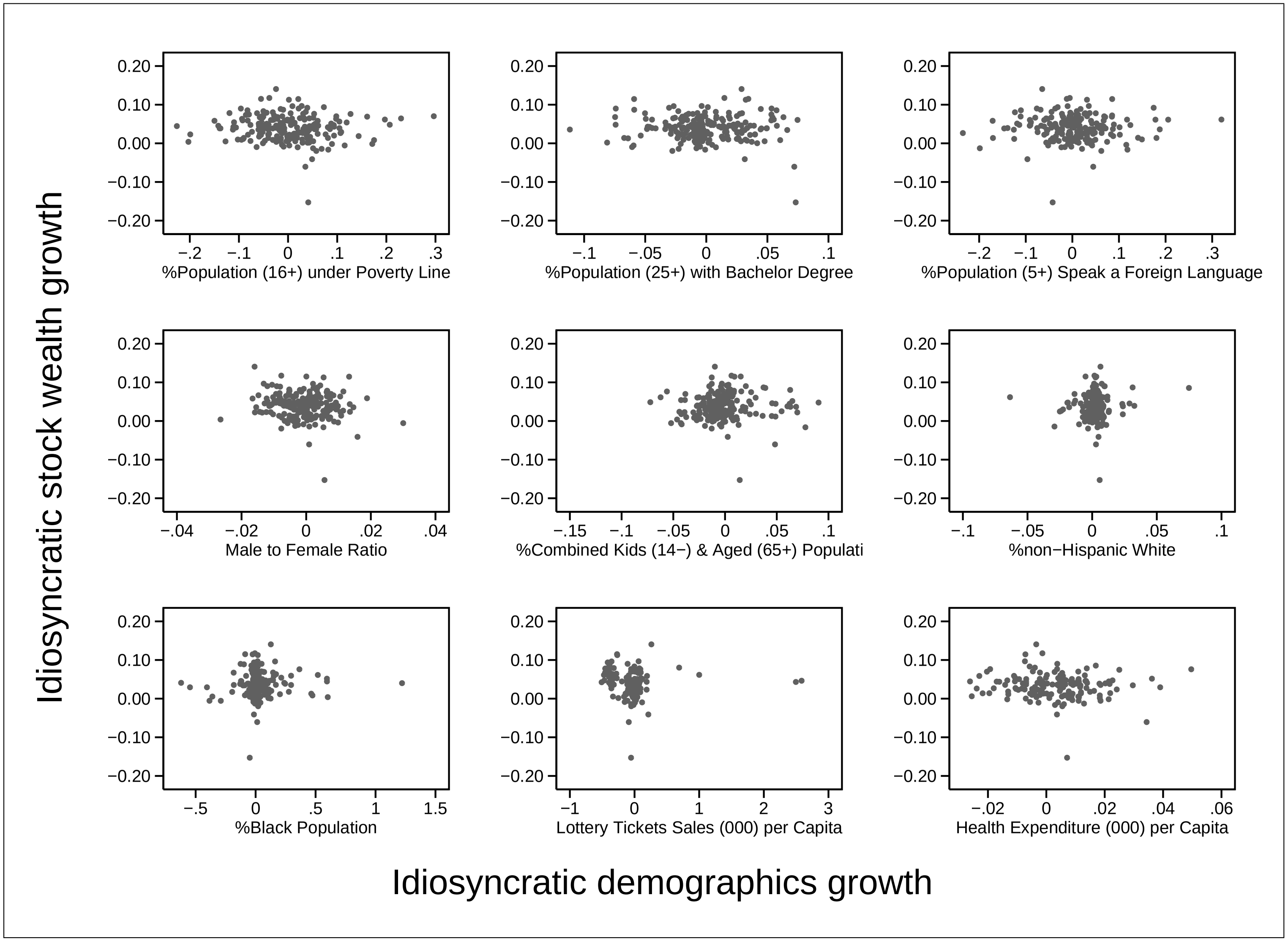


Figure 2.9: Growth of stock wealth versus demographic levels

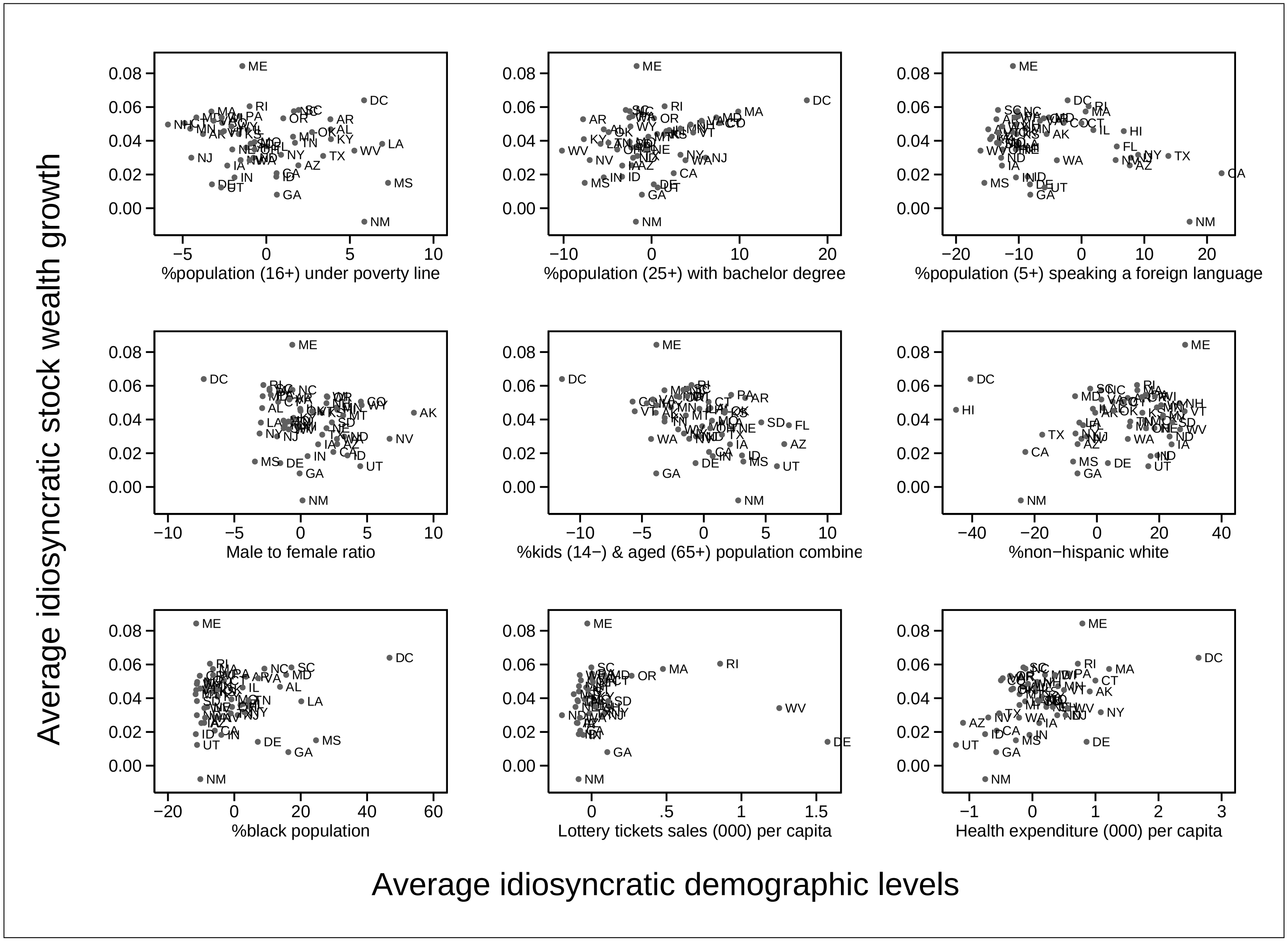


Figure 2.10: Standard deviation of stock wealth growth versus demographic levels

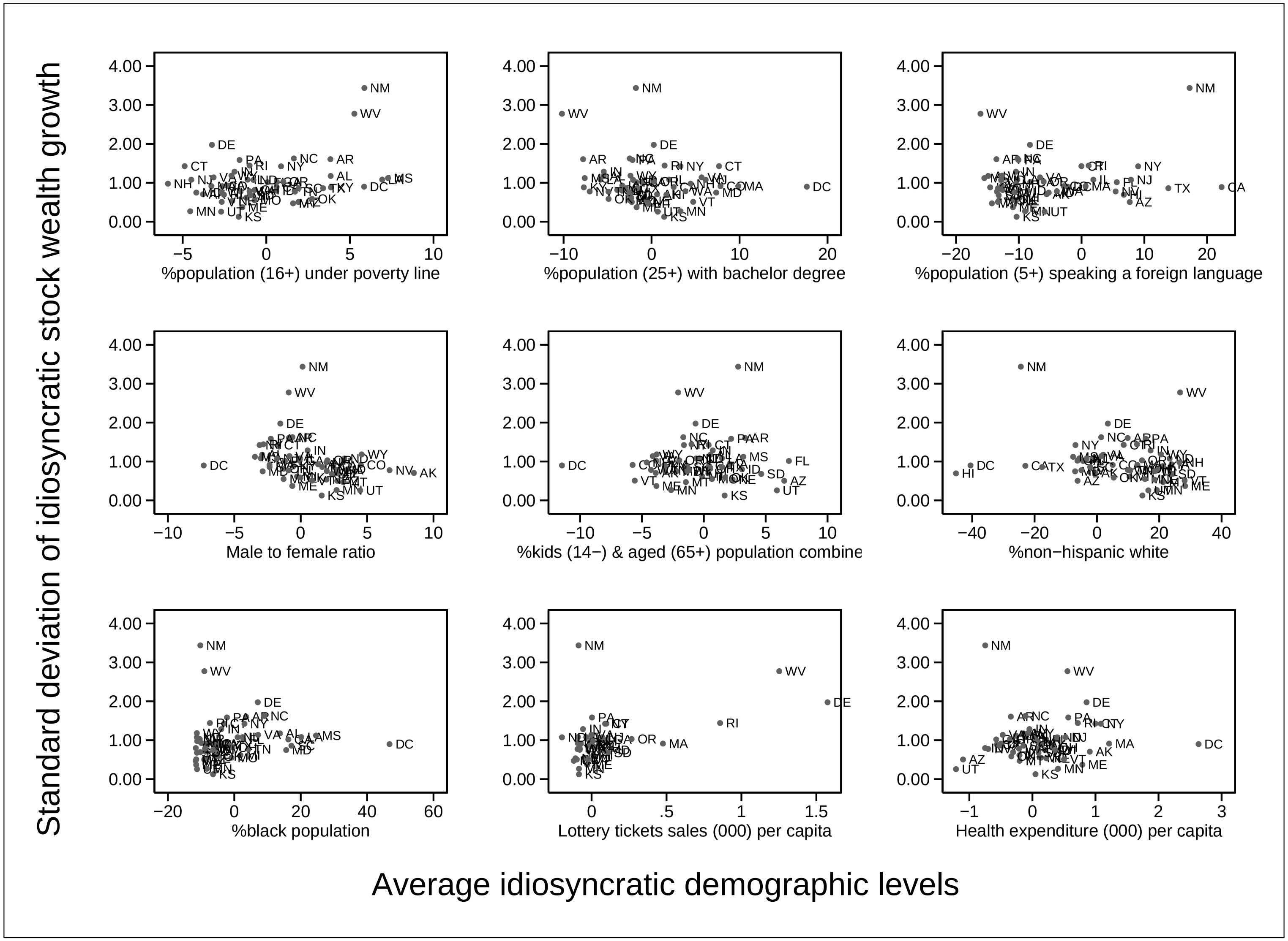


Table 2.1: IXI versus the SCF and FFA for 2001 (in trillions of dollars)

	SCF 2001	FFA 2001	IXI 2001h2[17]	IXI/SCF	IXI/FFA
Stocks	6.56	6.72	2.15	32.8%	31.9%
Mutual Funds	3.10	2.83	1.59	51.2%	56.2%
Total	9.66	9.56	3.74	38.7%	39.1%

Table 2.2: F-Test of the regional effect: $\beta_J = 0$ for $J = 2, ..., 9$

	$\Delta\tilde{w}^f_{i,t} = \alpha + \sum_{J=2}^{9} \beta_J * (i \in J) + \varepsilon_{i,t}$	$\Delta\tilde{w}^f_{i,t} = \alpha_t + \sum_{J=2}^{9} \beta_J * (i \in J) + \varepsilon_{i,t}$
F-Stat	0.604	0.954
d.f.	$(8, 501)$	$(8, 492)$
p-val	0.775	0.471

Table 2.3: $Std(\Delta\tilde{w}^f_{i,\cdot}) = \alpha + \beta * SPEC_i + \gamma * \overline{\Delta}_i \tilde{y}^f + \delta * \sigma(\Delta\tilde{y}_{i,\cdot})$

	All States			Washington, D.C. excluded		
β	0.005	0.002	0.006	0.006	0.003	0.006
	(0.005)	(0.007)	(0.006)	(0.009)	(0.01)	(0.009)
γ		0.471			0.481	
		(0.603)			(0.611)	
δ			-.181			-.180
			(0.347)			(0.368)
Obs.	51	51	51	50	50	50
$\bar{R}^2$	-.001	-.009	-.016	-.01	-.018	-.026

Table 2.4: Stock ownership and share of stock wealth by income quartile, SCF 2004

	Stock ownership	Share of total stock wealth held
Top quartile	57%	81%
Upper middle quartile	34%	12%
Lower middle quartile	19%	4%
Bottom quartile	9%	2%

Table 2.5: Correlation between stock wealth growth and demographic variables at the state level: 2001 – 2005

Demographic Variables[a]	$\rho(\Delta\tilde{w}^f_{i,t}, \Delta\tilde{\blacksquare}_{i,t})$[b]	$\rho(\bar{\Delta}_i\tilde{w}^f, \bar{\tilde{\blacksquare}}_i)$[c]	$\rho(\sigma_i(\Delta\tilde{w}^f), \bar{\tilde{\blacksquare}}_i)$[d]
%poverty	-0.054 (0.07)	-0.196 (0.14)	0.349** (0.134)
%bachelor degree	-0.063 (0.07)	0.297** (0.136)	-0.184 (0.14)
%home language not English	0.01 (0.07)	-0.268 * (0.138)	0.188 (0.14)
Sex-ratio	-0.123* (0.07)	-0.223 (0.139)	-0.297** (0.136)
Age-dependency	0.023 (0.07)	-0.445*** (0.128)	0.039 (0.143)
%white	0.076 (0.07)	0.175 (0.141)	-0.173 (0.141)
%black	0.046 (0.07)	0.092 (0.142)	0.092 (0.142)
Lottery	-0.089 (0.082)	-0.028 (0.162)	0.502*** (0.14)
Health expenditure	-0.083 (0.081)	0.465***[e] (0.126)	0.075 (0.142)

[a]Please refer to Section 2.6 for more detailed definition of the variables.
[b]$\Delta\tilde{w}^f_{i,t} = \Delta w^f_{i,t} - \Delta w^f_{\text{U.S.},t}$.
[c] $\bar{\Delta}_i\tilde{w}^f = \Sigma_{t=2001}^{2005}\Delta\tilde{w}^f_{i,t}/5$; $\bar{\tilde{\blacksquare}}_i = \Sigma_{t=2001}^{2005}(\blacksquare_{i,t} - \blacksquare_{\text{U.S.},t})/5$.
[d]$\sigma_i(\Delta\tilde{w}^f)$ is normalized by $\Sigma_{i=1}^{51}\sigma_i(\Delta\tilde{w}^f)/45$.
[e]

1. A five year period is too short to draw conclusions about the long-term relationship between demographic variables and wealth growth. (Please recall that these results only summarize the period of 2001 – 2005.)

2. The final result should be interpreted as indicating that in states where the level of health expenditures over the sample period was high, the growth rate of wealth tended to be high.

Table 2.6: Summary statistics of demographic variables

Demographics	$\tilde{\blacksquare}_i$					$\Delta\tilde{\blacksquare}_i$				
	Count	mean	std	min	max	Count	mean	std	min	max
%Poverty	306	-0.218	3.223	-6.9	8.5	255	0.003	0.076	-0.226	0.297
%Bachelor Degree	306	-0.346	5.138	-10.7	20.7	255	-0.001	0.032	-0.112	0.075
%Home Language not English	306	-5.759	9.046	-17.2	22.9	255	0.003	0.068	-0.235	0.320
Sex-Ratio	306	0.423	2.877	-7.981	9.525	255	-0.000	0.007	-0.026	0.030
Age-Dependency	306	-0.555	3.573	-12.668	9.142	255	-0.003	0.030	-0.123	0.090
%White	306	6.215	16.349	-46.748	29.2	255	0.003	0.010	-0.064	0.075
%Black	306	-1.306	11.548	-11.9	48.818	255	0.022	0.171	-0.657	1.221
Lottery	228	0.095	0.718	-0.198	7.885	188	-0.040	0.328	-0.508	2.585
Health Expenditure	255	0.083	0.654	-1.311	3.012	204	0.003	0.013	-0.029	0.056

Chapter 3

Constructing a New Measure of U.S. State-Level Spending

3.1 Introduction

Running at about 70 percent of U.S. GDP according to the Bureau of Economic Analysis, consumption is its largest component. Additionally, it constitutes about 85 percent of aggregate personal income and 97 percent of aggregate disposable personal income. Despite the apparent importance of consumption as reflected in the numerous studies on consumption behavior, few researchers have elaborated at length about the consumption data used to support their conclusions.

To date, the majority of empirical studies on consumption have employed aggregate or household-level data, most likely because these are the most readily available sources. However, studies using aggregate data are subject to endogeneity and aggregation problems, while household-level data suffers from serious measurement error problems.[1] Nevertheless, there exists an alternative option, one that might avoid some of the problems related to both macro and micro data – regional-level data. There are, however, very few studies in the current literature utilizing regional variations. The reason would at first seem discouraging – in the U.S., there is no measure of consumption at any level lower than the national level.

[1] See discussions in the first chapter of this dissertation.

Researchers, therefore, have often resorted to retail sales, which would be a valid approach if the focus were on growth rather than the level of consumption.

Although several sets of retail sales measures are available for U.S. states, there is no consensus in the current literature regarding which one is the best measure, neither regarding the strengths and weaknesses of any given dataset. Different researchers have used different data sets without explaining the reasons for their respective choice, most likely because they were unaware that alternatives existed.

This chapter presents an updated and improved version of the data derived from one of the sources found in the current literature; it then compares all of the sources, both with one another and with the aggregate data. This serves two purposes: (1) it provides supporting evidence for the choice of retail sales measures used in Chapter 4 of this dissertation; and (2) it gives future researchers a sense of each dataset's relative strengths and weaknesses, thus helping them decide which one best suits their respective needs.[2] More specifically, this chapter focuses on specific datasets such as were used in three different studies – Asdrubali, Sorensen, and Yosha (1996); Garrett, Hernàndez-Murillo, and Owyang (2004); and this dissertation. This chapter is organized as follows: Section 2 lists and briefly describes all of the retail sales datasets currently available for U.S. states. Section 3 discusses at length the strengths and weaknesses of each dataset in turn, and compares them at the national and state levels. Section 4 concludes.

3.2 Available datasets

Five different retail sales datasets for U.S. states can be found in the existing literature. Table 3.1 lists them together with the studies employing them. The data comprising the first dataset C^{HS}, first used by Hess and Shin (1998), is reported in the Census Bureau's Monthly Retail Trade Survey (RTS). RTS is probably the single most important source for estimates of national Personal Consumption Expenditures (PCE), which are reported by

[2] Discussion relies on the availability of the information revealed for each dataset, and is to the author's best understanding of the information.

the Bureau of Economic Analysis (BEA).[3] Furthermore, C^{HS} is the only direct measure of retail sales provided by a government statistical agency. Consequently, it is considered to be the most accurate. Therefore, it is used as a benchmark against which the qualities of the other datasets are measured. However, it is only available for 19 large states;[4] and moreover, it was discontinued in January 1997.

The second dataset, C^{SMM}, is prepared by the private company Claritas, and is published as the Survey of Buying Power in the magazine *Sales & Marketing Management*. It was first used by Asdrubali, Sorensen, and Yosha (1996), and was subsequently used in several other studies. By far, it is the most utilized dataset of state-level retail sales measures in the recent literature. Additionally, it is published in the Domestic Trade Section of the Census Bureau's *Statistical Abstract of the United States*. However, little information is provided such as would justify its popularity and seeming authority. Therefore, C^{SMM} is one of the focuses of this study. A detailed description of this dataset is given in the next section.

The third dataset, C^{CQS}, refers to the retail sales data used by Case, Quigley, and Shiller (2005), whose study is particularly relevant to the fourth chapter of this dissertation. A brief description of the data is provided in Case, Quigley, and Shiller (2005), and is quoted below:

> [A] consistent panel of retail sales has been constructed by Regional Financial Associates (RFA) ...... The RFA estimates were constructed from county level sales tax data, the Census of Retail Trade (CRT) published by the U.S. Census Bureau, and the Census Bureaus monthly national retail sales estimates. For states with no retail sales tax or where data were insufficient to support imputations, RFA based its estimates on the historical relationship between retail sales and retail employment. Data on retail employment by state are available from the Bureau of Labor Statistics. Regression estimates relating sales to employ-

[3] See Wilcox (1992).

[4] They are California, Florida, Illinois, Indiana, Louisiana, Massachusetts, Maryland, Michigan, Minnesota, Missouri, North Carolina, New Jersey, New York, Ohio, Pennsylvania, Texas, Virginia and Wisconsin.

> ment were bench-marked to the Census of Retail Trade available at five-year intervals. Estimates for all states were within five percent of the benchmarks.[5]

Apart from the above quotation, no further information regarding the data construction is provided. Given its clear dependency on employment only, this paper does not focus on C^{CQS}. However, it will be investigated in the Appendix, where the relationship between retail sales measures and state employment is explored.

C^{GHO} was first utilized by Garrett, Hernàndez-Murillo, and Owyang (2004). It was calculated using state sales tax revenue collections and state retail sales tax rates. C^{ZHOU} uses the same data resource for most states and time periods, though whenever possible, incorporates more accurate measures of retail sales from state tax agencies. More information can be found in the next section.

3.3 Validity check of available datasets

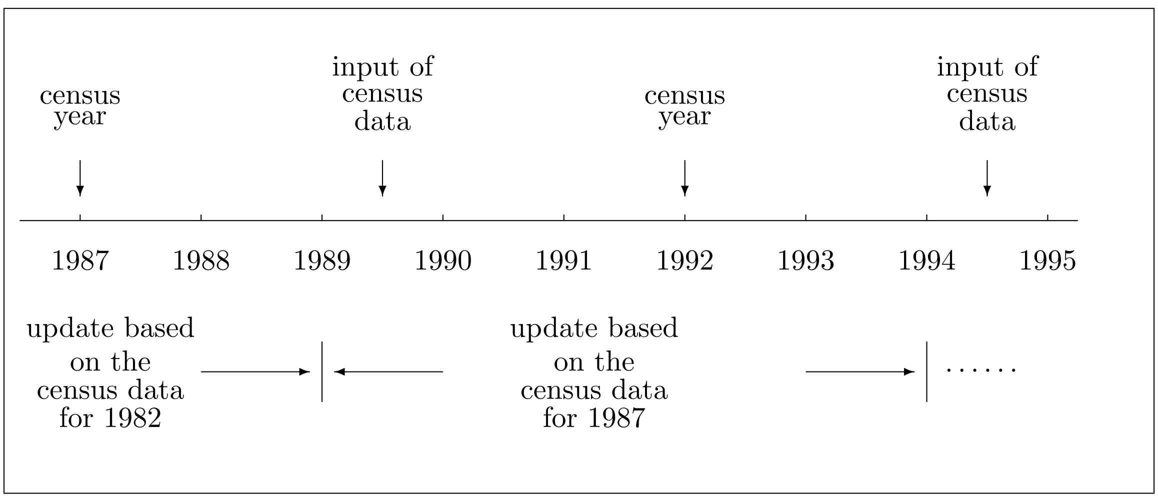
This section examines the validity of C^{SMM}, C^{GHO}, and C^{ZHOU}. We first describe the construction of each of the three datasets, and then discuss which one is likely a better measure of consumption according to their constructions. Additionally, in order to understand their performances in measuring consumption growth, we compare all three datasets with the benchmarks at both the aggregate and the state level.

3.3.1 Construction of C^{SMM}

C^{SMM} is available annually for all 50 states and the District of Columbia, as well as all metro areas and certain large cities and counties. Total retail sales series are broken down into five major categories by store type and merchandize line through 1999 according to the Standard Industrial Classification (SIC), and since then, by the North American Industry Classification System (NAICS).

[5] RFA is now a subdivision of Moody's Economy.com.

Based on the author's personal communications with the company Claritas and corporative online descriptions,[6] it appears that Claritas updates its retail sales estimates using the most recent *Census of Retail Trade*, rather than its own most recent *estimates* of retail sales. Specifically, annualized rates of change from the most recent *Census of Retail Trade* at the regional, state and county levels are calculated using local sales tax data, together with wage and employment data (the EOS-202 file) from the Bureau of Labor Statistics (BLS). Finally, Monthly Retail Trade data at the aggregate level are used as control totals for the purpose of adjusting the estimates at lower geographical levels. Through 2000, the survey provided previous calendar year estimates, but since 2000, it began reporting estimates for the current year. Consequently, the retail sales estimate for 1999 is missing. The following is an implied time line the author based on Claritas' documentation.[7]



However, Claritas provides no further detailed information regarding how the annualized growth rate is constructed.

3.3.2 Construction of $C^{\mathbf{GHO}}$ and $C^{\mathbf{ZHOU}}$

Garrett, Hernàndez-Murillo, and Owyang (2004) introduced a new method for constructing taxable retail sales data. They divide state general retail sales tax revenue by the state sales tax rate for the corresponding time period. Quarterly state sales tax series[8] are published

[6] http://www.claritas.com/claritas/Default.jsp?ci=3&si=1&pn=retailmktpower.

[7] One which may not reflect the reality.

[8] The general retail sales tax receipts collected by state governments only. These are not equivalent to the total tax revenues collected by state and local governments.

in the U.S. Census Bureau's *State Government Tax Collections*; state sales tax rates can be obtained from various sources such as the *State Government Tax Collections* and the Tax Foundation's *Facts and Figures on Government Finances*. Since this calculation requires sales tax information, the imputed C^{GHO} is only available for 45 states plus the District of Columbia.[9] Furthermore, the sales tax revenue data for Nevada has been dropped, on account of its being incomplete and highly volatile.

Changes in C^{GHO} are expected to be volatile and sometimes significantly different from those in retail sales, as sales tax collection varies with tax policy. For example, a state might exempt a certain type of goods from the sales tax for a certain period of time. A major part of the subsequent effect on that state's sales tax collection, and hence calculated retail sales, is thus independent of real retail sales. Nevertheless, this effect can hardly be identified, much less separated from retail sales variations.

C^{ZHOU} improves C^{GHO} by incorporating the taxable retail sales or gross retail sales data published by state tax agencies. Taxable sales or gross retail sales numbers might come directly from the actual figures reported on dealers' returns (as in Iowa), or be computed based on sales tax revenues collected by state tax agencies (as in Virginia). In the former case, the reported taxable or gross retail sales numbers are, by construction, more accurate than the indirectly calculated C^{GHO}. Even in the latter case, the taxable retail sales computed by state tax agencies are expected to be much more accurate than C^{GHO} because of various reasons. First of all, state tax agencies have access to much richer information. For example, many states apply lower sales tax rates to food and drugs, but the data on sales tax revenues collected from these merchandise are not available to us.[10] Furthermore, certain types of consumer goods, such as cigarettes and liquor, among others, are included in special sales tax programs, where different sales tax rates apply. These revenues, therefore, are not included among the general retail sales receipts, which in turn makes our calculation of taxable retail sales less than comprehensive. Additionally, as policymakers

[9] States that do not have sales tax are Alaska, Delaware, Montana, New Hampshire and Oregon.

[10] Recent state sales tax rates for general sales, food items, prescription drugs and non-prescription drugs are available at http://www.taxadmin.org/FTA/rate/sales.html

and data collectors, state tax agencies are more likely to capture data errors, and thus be able to correct them.[11] However, despite this data source's superiority, state government reported taxable or gross retail sales data is only available for 12 states, and in all cases, only for a limited period of time. Furthermore, the frequency of the data varies by state, ranging from monthly to annually. Nevertheless, in our study, C^{GHO} and C^{ZHOU} are both converted into annual frequencies for comparison purposes.

Since C^{ZHOU} makes use of various data sources, its measure of retail sales is not comparable over time or across states. Taking Florida as an example, through 1994, C^{ZHOU} measures the same taxable retail sales as C^{GHO} does. From 1995 onward, however, C^{ZHOU} records the government reported gross retail sales. To construct a consistent measure of retail sales, we start by constructing a consistent set of retail sales growth by taking the log differences of the sales data from the same source. C^{ZHOU} is then reconstructed from 1970 onward using the C^{GHO}. Whenever available, the growth in government reported taxable or gross retail sales is utilized. Otherwise, $\Delta C^{\mathrm{GHO}}_{i,t}$ is used. The specific formulas are given below:

1. $\Delta C^{\mathrm{ZHOU}}_{i,t} = \Delta RetailSales_{i,t}$, when $\Delta RetailSales_{i,t}$ is available
2. $\Delta C^{\mathrm{ZHOU}}_{i,t} = \Delta TaxableSales_{i,t}$, when $\Delta TaxableSales_{i,t}$ is available
3. Otherwise, $\Delta C^{\mathrm{ZHOU}}_{i,t} = \Delta C^{\mathrm{GHO}}_{i,t}$.
4. $C^{\mathrm{ZHOU}}_{i,t=1970} = C^{\mathrm{GHO}}_{i,t=1970}$
5. $C^{\mathrm{ZHOU}}_{i,t+1} = C^{\mathrm{ZHOU}}_{i,t} * \exp \Delta C^{\mathrm{ZHOU}}_{i,t}$ for t =1970 through 2005.

3.3.3 Data cleaning

Retail sales and consumer spending fluctuate in response to tax rate movements or tax policy changes. However, because of data limitations, it is difficult to distinguish this fluctuation from measurement errors. For example, new tax rates might be implemented in the middle of a quarter. Therefore, our imputed quarterly retail sales measure would be biased

[11] Some comparisons are given in the next section.

in the opposite direction from the tax rate changes. Another example would be a drop in retail sales taxes, and therefore the imputed retail sales measures due to temporary tax rate exemption periods. Both $\Delta C_{i,t}^{\text{GHO}}$ and $\Delta C_{i,t}^{\text{ZHOU}}$ are subject to potential temporary shocks that are unrelated to real retail sales and consumption measures. Figure 3.3 shows how volatile $\Delta C_{i,t}^{\text{GHO}}$ and $\Delta C_{i,t}^{\text{ZHOU}}$ can be during periods when the tax rate changes. It provides evidence of the need to eliminate variations that are independent of retail sales.

This paper implements two methods to identify fluctuations in $\Delta C_{i,t}^{\text{ZHOU}}$ that are potentially irrelevant with respect to real retail sales or consumer spending.[12] As a first step, $\Delta C_{i,t}^{\text{ZHOU}}$ is dropped for periods where the tax rate changes are equal to or greater than 1 percent. Any remaining $\Delta C_{i,t}^{\text{ZHOU}}$ that are greater by 3 times the standard error than the mean for each state are also dropped, as they are considered outliers. Altogether, 88 outliers are identified and dropped. Finally, all missing $\Delta C_{i,t}^{\text{ZHOU}}$ are replaced with the average $\Delta C_{i,t}^{\text{ZHOU}}$[13] for each state. The new series of the growth rate and the level of the retail sales measure are then notated as $\Delta C_{i,t}^{\text{ZHOUG}}$ and $C_{i,t}^{\text{ZHOUG}}$ respectively.[14]

3.3.4 Validity of C^{SMM}, C^{GHO}, and C^{ZHOU} as a consumption measure

This section discusses that, based on the construction of C^{SMM}, C^{GHO}, and C^{ZHOU}, which one would act as a better measure of consumption. We believe that the answer depends on which question the data is being used for because each dataset has its own strengths and weaknesses.

Based on the limited information available, C^{SMM} might be a better choice for those more concerned with the level of retail sales or data coverage. Some major reasons are listed below:

- C^{SMM} is the estimate of total retail sales, while C^{GHO} and C^{ZHOU} only reflect taxable

[12] State tax agency provided retail sales measures are assumed to be affected to a much lesser degree by the above-mentioned measurement errors, and therefore are not subject to the following outlier identification practice.

[13] The average $\Delta C_{i,t}^{\text{ZHOU}}$ is re-calculated after the outliers are dropped.

[14] The data archive that can produce all results in this study is available from the Johns Hopkins library, at URL: http://jhir.library.jhu.edu/handle/1774.2/34267.

retail sales.

- C^{SMM} is available for all states, while C^{GHO} and C^{ZHOU} are only available for 44 states plus the District of Columbia.
- C^{SMM} is also available for many lower geographic levels, such as metro areas, counties, and cities.
- C^{SMM} is smoother than C^{GHO} and C^{ZHOU}, since some variations found in C^{GHO} and C^{ZHOU} are not caused by changes in retail sales, but rather by changes in tax policies.

However, researchers concerned about the growth of retail sales or who are looking for the true relationship between consumption and other variables, like income, might arrive at a different answer.

- The construction of C^{SMM} depends on the *estimates* for ΔC^{SMM}. Although the exact formula or model used to estimate ΔC^{SMM} is proprietary, we do know that state employment and wages are incorporated in the estimation. This practice may make a lot of sense to those who care about market shares or market potentials. It is inappropriate, however, for studies investigating income effects, wealth effects, risk-sharing analyses, and so forth, as the conclusions drawn from the estimates of ΔC^{SMM} depend to a large extent on how employment and wages are used in constructing ΔC^{SMM}. To give a simple example, let us assume that Claritas uses wages only to estimate the growth in retail sales:

$$\Delta \hat{C}^{\text{SMM}} = \hat{\alpha} + \hat{\beta} * \Delta W,$$

  where changes in real retail sales are actually determined by ΔW, another variable ΔX, and an error term, ε, reflective of preference shocks:

$$\Delta C = \alpha + \beta * \Delta W + \gamma * \Delta X + \varepsilon.$$

  Without losing generality, the change in income, ΔY, is assumed to be identical to the change in ΔW. Based on the construction, the real income effect should be β. The income effect estimated using ΔC^{SMM}, however, will be exactly $\hat{\beta}$. In other words, no matter how hard a researcher try, the best possible estimate of income effect will depend on a coefficient that is assumed in advance and is neither documented nor explained.

- Each set of estimates of C^{SMM} uses data that goes back to the most recent census on retail trade. The main benefit of this practice is that it avoids accumulating estimation errors or biases, while incorporating more information over time. This practice gives rise to smoothed retail sales series. On the other hand, it means that the difference between the respective estimates of two consecutive years does not represent the growth in retail sales over those two years. Quoted below is a short paragraph discussing trend-related issues taken from Claritas' demographic methodology documentation.[15] It should apply equally to Claritas' construction of retail sales.

  > To take full advantage of methodological refinements and new data resources, each set of updates begins not with the previous year's estimates, but with data from the most recent decennial census. For this reason, the difference between estimates for consecutive years is not an estimate of change from one year to the next. Change is estimated with reference to the previous census numbers.

  This paper attempts to reverse the engineering of C^{SMM}, by investigating how it is related to state employment, wages, and national retail sales. Other measures of consumption are also studied for comparison purposes. Please refer to the Appendix for detailed discussions and results.

- Claritas has reported current year estimates for retail sales since 2000. This change implies that any shock to retail sales occurring in the current year, but after data publication, will not be reflected in the current year estimates. Of Course, large impacts of well-known events such as Hurricane Katrina could cause Claritas to produce a special report after all data reflecting the impact becomes available. Such data

[15] Discussion of trend-related issues in Claritas' documentation can be found online: http://www.claritas.com/claritas/demographics.

corrections are unlikely to take place, however, for region-specific and less important shocks. Therefore, this change might be problematic for researchers who need the most accurate measures.

3.3.5 Comparison with the benchmark at the aggregate level

This section investigates the validity of the datasets by comparing them with the benchmark measure at the aggregate level. Specifically, we study whether the sum of C^{SMM}, C^{GHO}, C^{ZHOU} and C^{ZHOUG} might act as a good proxy for U.S. aggregate retail sales growth rate. Likewise, it also investigates whether one measure is significantly better than the others at the national level.

Figure 3.1 is a plot of $\Delta \sum c^{\mathrm{SMM}}$[16] and $\Delta \sum c^{\mathrm{ZHOUG}}$ versus $\Delta retail^{\mathrm{U.S.}}$. To avoid confusion, $\Delta \sum c^{\mathrm{ZHOU}}$ is not presented, since it is identical to $\Delta \sum c^{\mathrm{ZHOUG}}$ for most time periods at the aggregate level, and therefore would be indistinguishable from $\Delta \sum c^{\mathrm{ZHOUG}}$ in the figure. Figure 3.1 shows that all three series do a decent job in presenting aggregate retail sales growths.

Table 3.2 reports if one unit change in $\Delta retail^{US}$ could be reflected in $\Delta \sum c^{\mathrm{SMM}}$, $\Delta \sum c^{\mathrm{GHO}}$, $\Delta \sum c^{\mathrm{ZHOU}}$, or $\Delta \sum c^{\mathrm{ZHOUG}}$. It shows that about 80 percent of the variations in real aggregate retail sales growth can be reflected in the growth of all four series, which indicates that they are all reasonable candidates for measuring retail sales as far as capturing growth in the aggregate term is concerned.

3.3.6 Comparison with the benchmark at the state level

It is more illuminating to compare C^{SMM}, C^{GHO}, C^{ZHOU}, and C^{ZHOUG} at the state level. Assuming that C^{HS} is the most accurate measure of retail sales at the state level, we can then explore the performances of C^{SMM}, C^{GHO}, C^{ZHOU} and C^{ZHOUG} by comparing them with C^{HS}. Of course, the comparison can only be done for those 19 states where C^{HS} is available.

[16] Lower case letters are used to denote corresponding real per capita terms.

As shown in Table 3.3, all four series other than ΔC^{ZHOUG} have average growth rates of around 0.3 percent, which is fairly low. Despite the fact that ΔC^{ZHOUG} shows improvements over C^{GHO} and C^{ZHOU} in terms of having less variation and more reasonable lower and upper bounds, at first sight, the average ΔC^{ZHOUG} seems surprisingly higher at 0.7 percent. On the other hand, the fact that most sizable tax rate changes are positive instead of negative – such that most excluded outliers are negative numbers – causes C^{ZHOUG} to be higher than the real numbers. The standard errors presented in Table 3.3 provide evidence that ΔC^{ZHOUG} is not statistically different from 0.3 percent. Furthermore, Figures 3.5 and 3.6, show that, apart from a few observations, ΔC^{ZHOUG} is almost identical to $\Delta c^{\text{HS}}_{i,t}$.

The correlations between $\Delta c^{\text{HS}}_{i,t}$ and $\Delta C^{\text{SMM}}_{i,t}$, $\Delta C^{\text{GHO}}_{i,t}$, $\Delta C^{\text{ZHOU}}_{i,t}$, and $\Delta C^{\text{ZHOUG}}_{i,t}$ are listed in Table 3.4. Generally speaking, Δc^{SMM} shows the greatest correlation with Δc^{HS}, while $\Delta C^{\text{ZHOU}}_{i,t}$ and $\Delta C^{\text{ZHOUG}}_{i,t}$ are more correlated with Δc^{HS} than $\Delta C^{\text{GHO}}_{i,t}$. One probable explanation for the former observation could be that C^{HS} is one of the sources Claritas uses to estimate Δc^{SMM}. The latter observation provides evidence for the superiority of $\Delta C^{\text{ZHOU}}_{i,t}$ over $\Delta C^{\text{GHO}}_{i,t}$.

It is more interesting to investigate if one unit change in $\Delta C^{\text{HS}}_{i,t}$ can be reflected in $\Delta C^{\text{SMM}}_{i,t}$, $\Delta C^{\text{GHO}}_{i,t}$, $\Delta C^{\text{ZHOU}}_{i,t}$, and $\Delta C^{\text{ZHOUG}}_{i,t}$. Table 3.5, however, shows that unity is not included in the 95 percent confidence interval for any of the four series. $C^{\text{ZHOU}}_{i,t}$ appears to be superior, in terms of having the highest coefficient and an upper bound of the 95% interval closest to unity. The differences between the four series, however, are not statistically significant.

So far, the above analysis has not demonstrated any dramatic differences between $C^{\text{GHO}}_{i,t}$ and $C^{\text{ZHOU}}_{i,t}$. This is because most state governments only very recently started reporting (if ever) retail sales or taxable sales. Therefore, few such replacements occurred between 1978 and 1996, the period during which C^{HS} was available. Virginia, however, is a unique case, in that its taxable sales data goes back to 1984. It provides a good opportunity for exploring whether Δc^{ZHOU} is a better measure of retail sales than Δc^{GHO}.

Still using Δc^{HS} as a benchmark, Table 3.6 shows that in the case of Virginia, Δc^{GHO} is a very poor substitute for ΔC^{HS}. In contrast, a one unit change in $\Delta C_{i,t}^{\mathrm{HS}}$ corresponds to an almost identical change in $\Delta C_{i,t}^{\mathrm{ZHOU}}$. This suggests that $C_{i,t}^{\mathrm{ZHOU}}$ constitutes a significant improvement over $\Delta C_{i,t}^{\mathrm{GHO}}$, and is a better measure of retail sales growth than $\Delta C_{i,t}^{\mathrm{SMM}}$. Figure 3.4 plots the four series together, and thus provides evidence reinforcing this conclusion.

3.4 Conclusion

This study compares 5 different retail sales measures. It shows that, for some purposes, C^{SMM} is the best measure. This is for various reasons: it is smooth, covers all U.S. states, has a certain degree of authority as it was adopted by the *Statistical Abstract*, and in general, has been widely used. For those who care about having consistent measures of the level of retail sales, C^{SMM} is probably the best choice. However, researchers who care about the growth of retail sales should be aware of several concerns. The fact that the growth rate is estimated, likewise that its variation can largely be explained by a simple linear model, raises concerns that C^{SMM} will likely reflect mainly the "model" behind the constructed data as opposed to the actual economic facts. C^{CQS} has the same concerns as C^{SMM}, and for some periods, presents worrisome dependencies on state employment, wage, and national retails sales. On the contrary, although C^{ZHOU} and C^{ZHOUG} are much less smooth and could not act as proxies for the level of retail sales, they are both good measures of retail sales growth. Where state taxable sales or retail sales measures are available, C^{ZHOU} and C^{ZHOUG} are shown to be significantly superior to C^{GHO}. Finally, the constructions of ΔC^{ZHOU} and $\Delta C^{\mathrm{ZHOUG}}$ do not involve any assumed models, such as might contaminate studies of consumption behaviors because of prior assumptions.

Figure 3.1: The sum of state retail sales measures versus U.S. aggregate retail sales

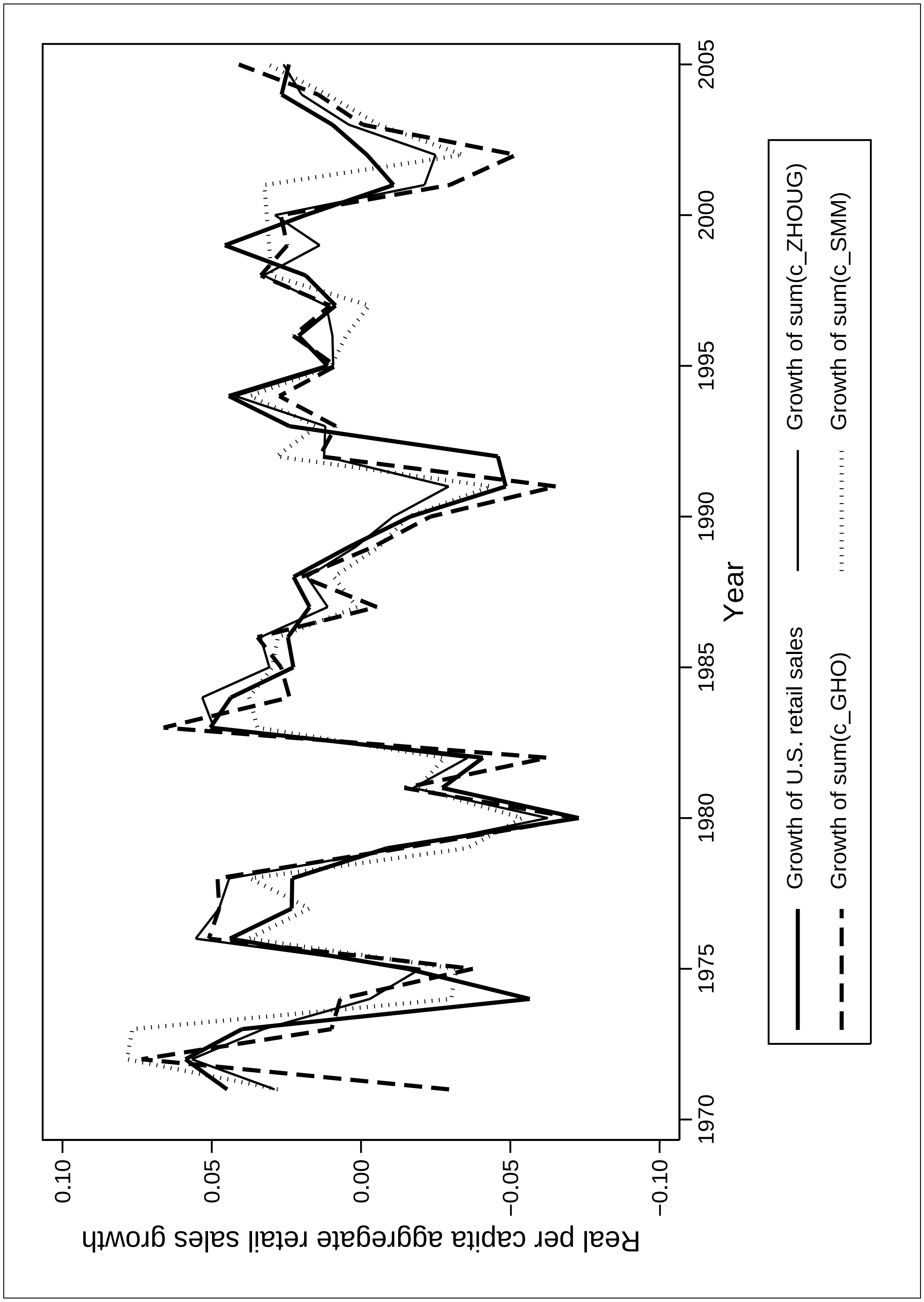


Figure 3.2: The sum of state retail sales measures versus U.S. aggregate consumption

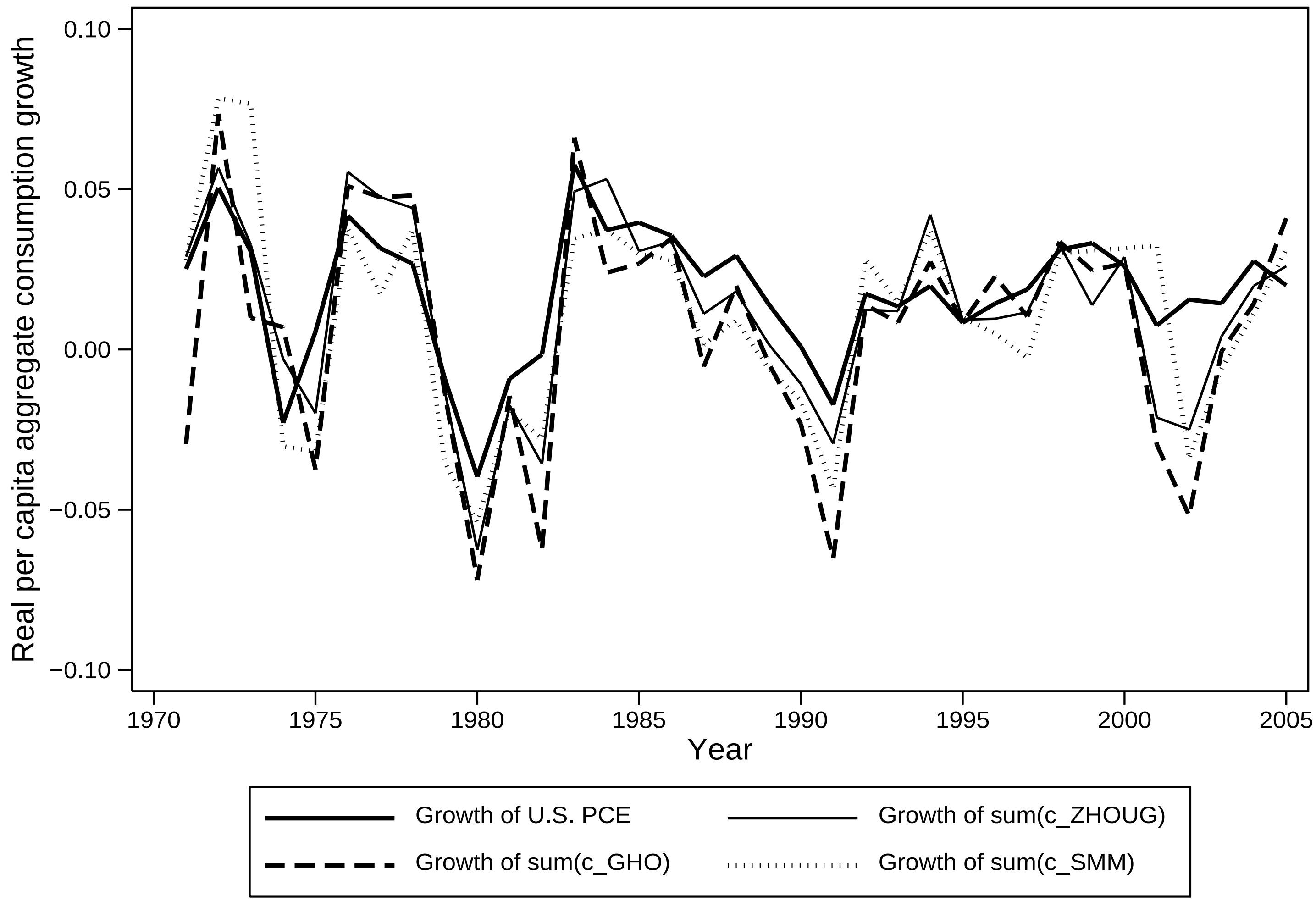


Figure 3.3: Texas: $\Delta c_{i,t}^{\mathrm{HS}}$ versus $\Delta c_{i,t}^{\mathrm{ZHOUG}}$, $\Delta c_{i,t}^{\mathrm{GHO}}$ and $\Delta c_{i,t}^{\mathrm{SMM}}$

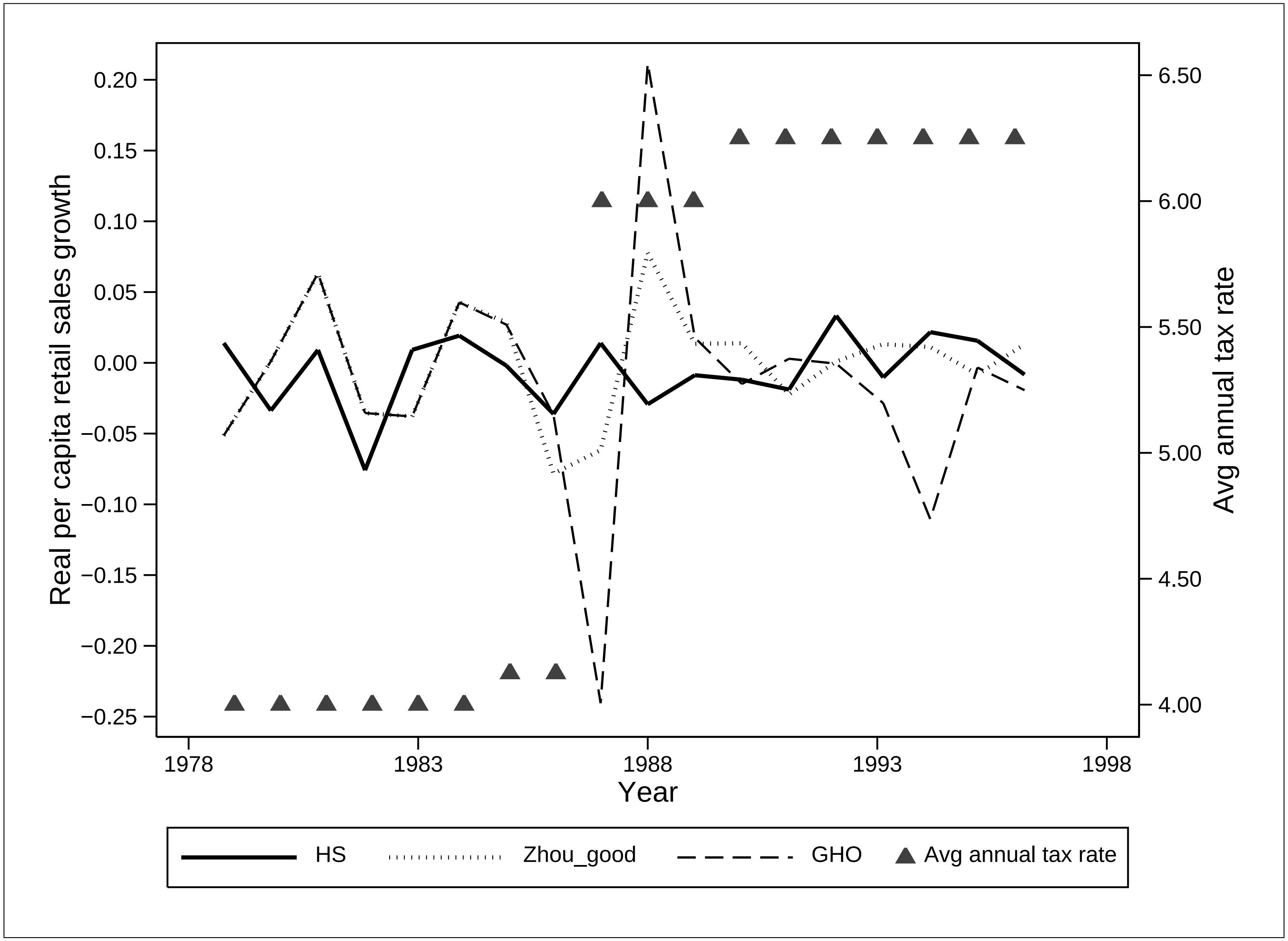


Figure 3.4: Virginia: $\Delta c_{i,t}^{\mathrm{HS}}$ versus $\Delta c_{i,t}^{\mathrm{ZHOU}}$, and $\Delta c_{i,t}^{\mathrm{GHO}}$

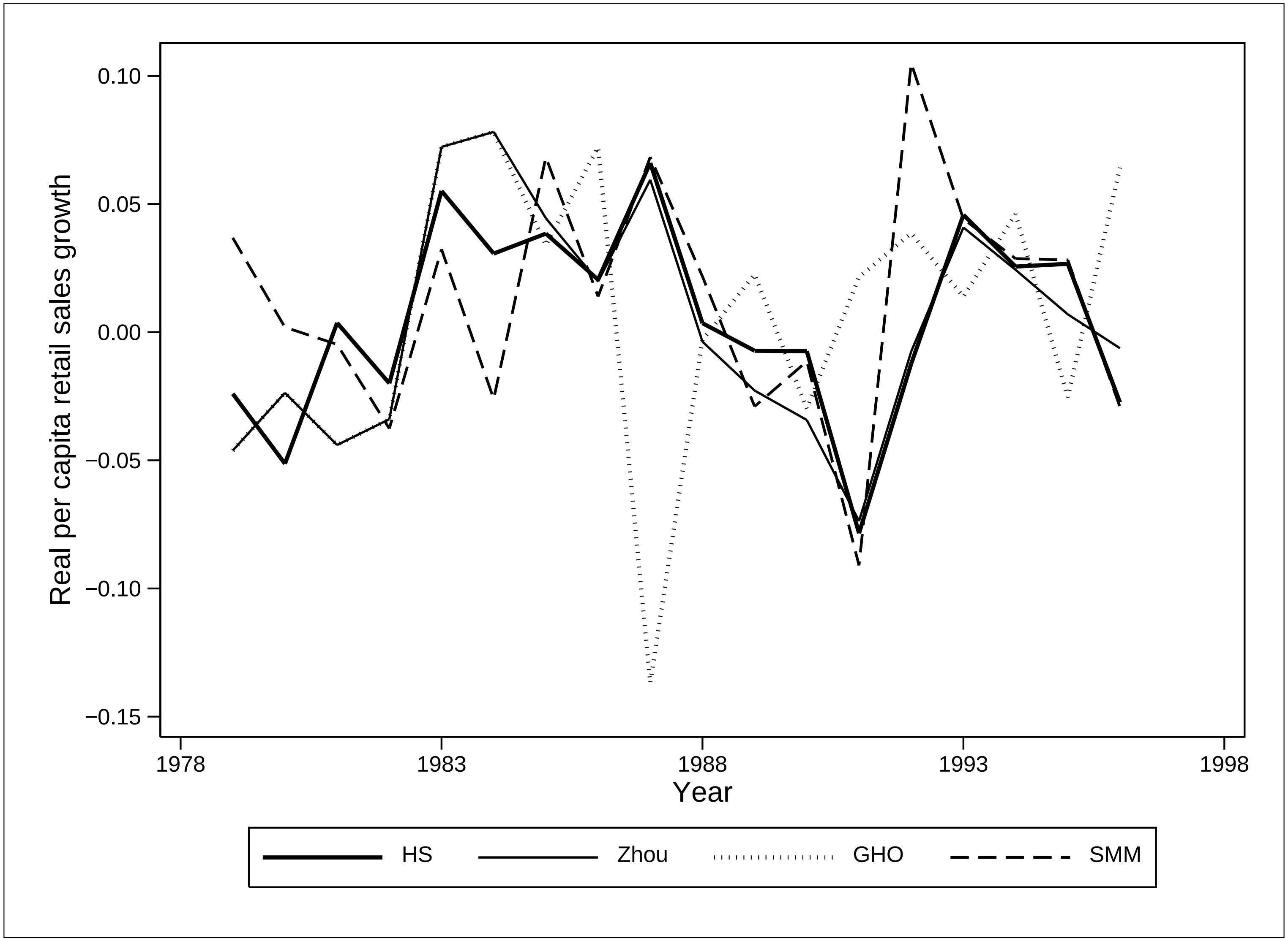


Figure 3.5: CA-MO: $\Delta c_{i,t}^{\mathrm{HS}}$ versus $\Delta c_{i,t}^{\mathrm{ZHOU}}$, and $\Delta c_{i,t}^{\mathrm{ZHOUG}}$

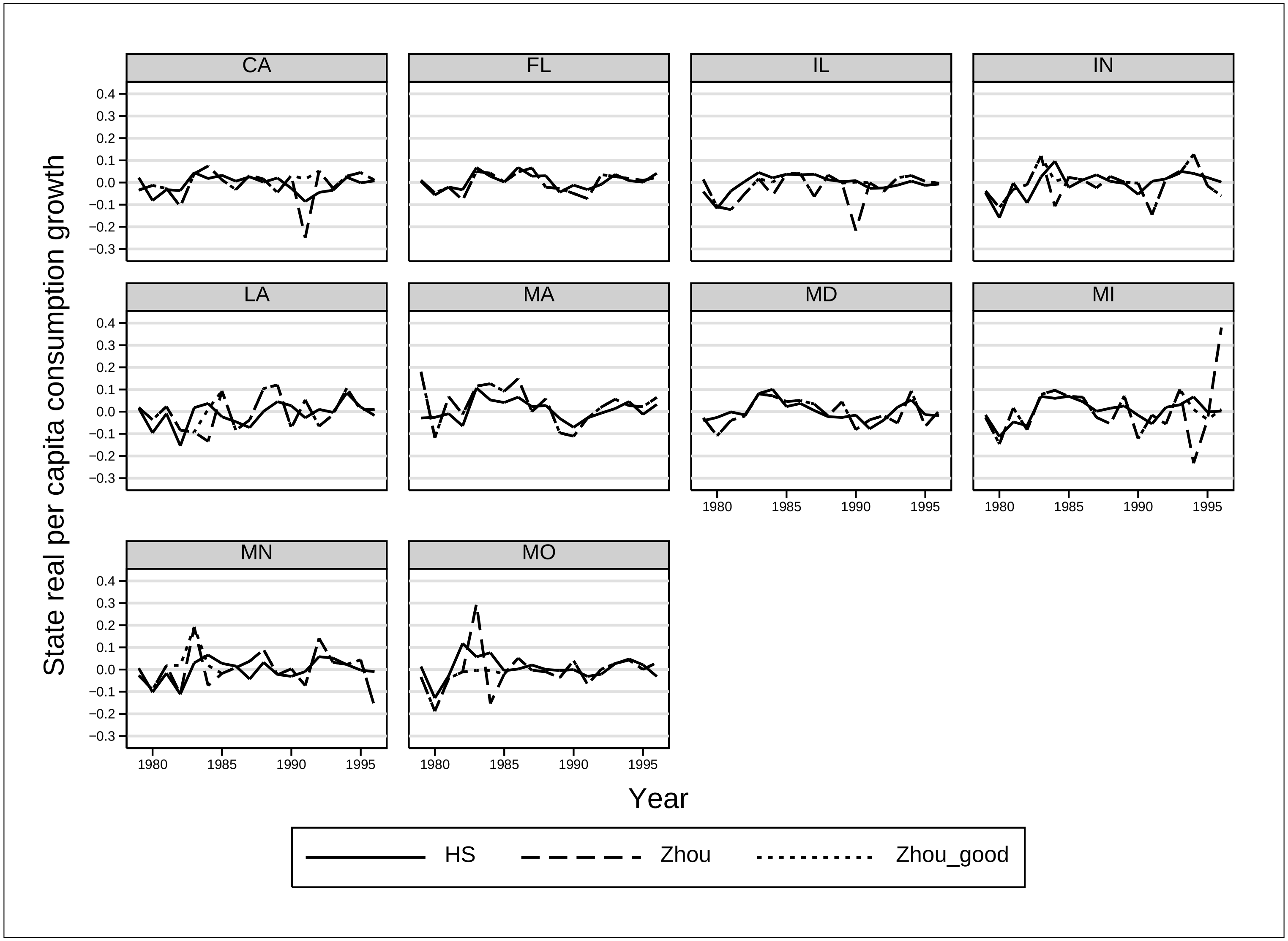


Figure 3.6: NC-WI: $\Delta c_{i,t}^{\mathrm{HS}}$ versus $\Delta c_{i,t}^{\mathrm{ZHOU}}$, and $\Delta c_{i,t}^{\mathrm{ZHOUG}}$

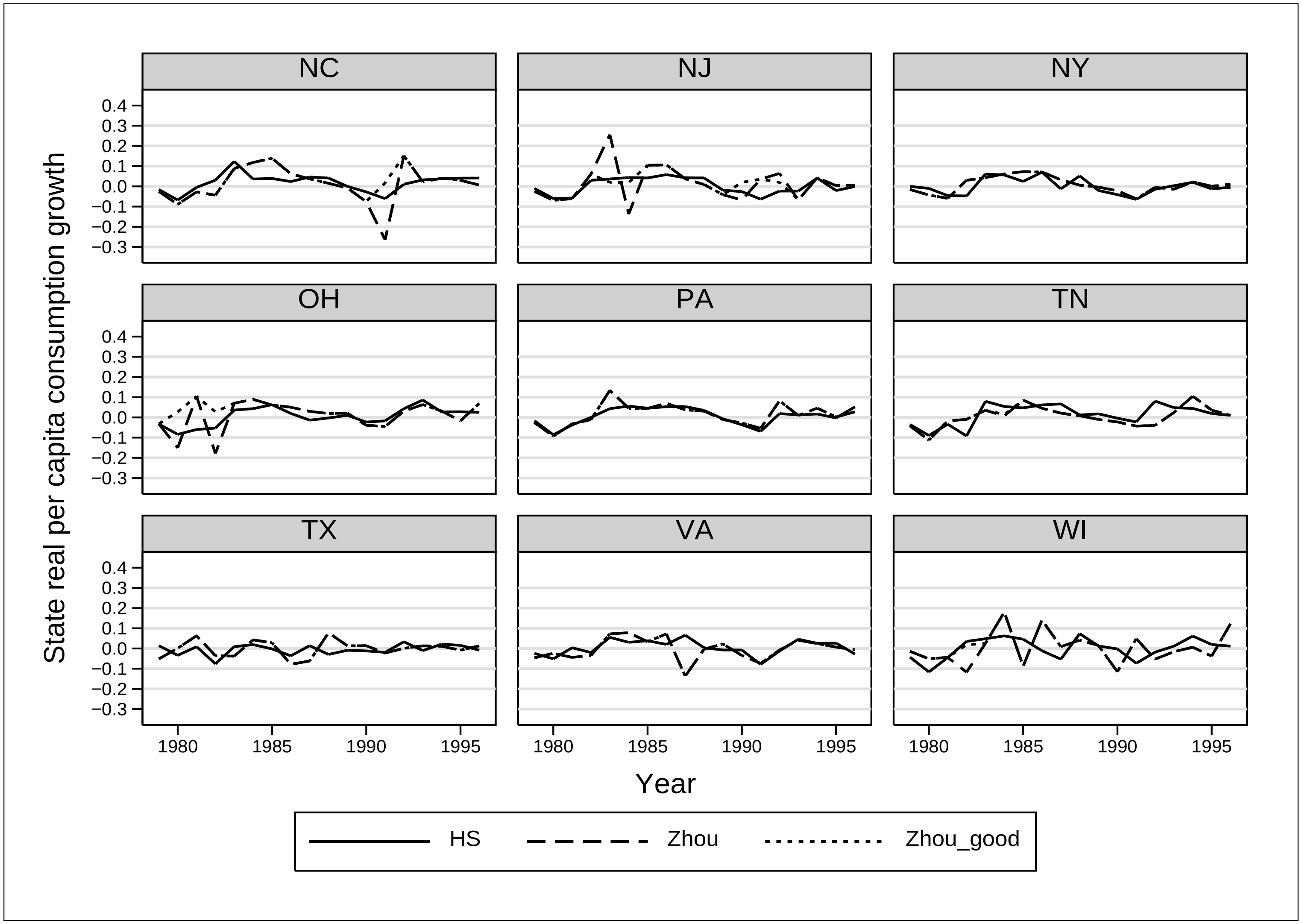


Table 3.1: Description of datasets

	Data sources	Works using data	Time range	States covered
C^{HS}	Monthly Retail Trade - Survey	Hess and Shin (1998) Del Negro (1998)	1978m01-1996m12	19 large states
C^{SMM}[a]	Sales & Marketing - Management	Asdrubali, Sorensen, and Yosha (1996) Del Negro (1998) Luengo-Prado and Sorensen (2006)	1963-1998 & 2000-present	51 states
C^{CQS}	Regional Financial Associates	Case, Quigley, and Shiller (2005)	1977Q3-2006Q4	51 states
C^{GHO}[b]	State Government - Sales Tax Collections	Garrett, Hernàndez-Murillo, and Owyang (2004)	1970q1-present	45 states
C^{ZHOU}, C^{ZHOUG}	C^{GHO} + taxable retail sales gross retail sales	Ravina (2005) Zhou(2010)	1970q1-present	45 states

[a]The data from 1963 to 1990 is downloadable from the late professor Oved Yosha's webpage http://econ.tau.ac.il/research/riskshare/channels/channels.htm. The data after 1990 are available in the Statistical Abstract.

[b]The data is downloadable from http://research.stlouisfed.org/publications/review/05/03/part1/0503rh.zip.

Table 3.2: $\Delta \sum_{i=1}^{51} c^*_{i,t} = \alpha_t + \beta \Delta retail_t^{US} + \varepsilon_t$ for 1971-1998, 2000-2005

$\Delta \sum_{i=1}^{51} c^*_{i,t}$	β	Std. Error	[95% Conf.	Interval]	$\bar{R}^2$	Obs
$\Delta \sum_{i=1}^{51} c^{\text{SMM}}_{i,t}$	0.76	0.11	(0.54	0.99)	0.60	33
$\Delta \sum_{i=1}^{51} c^{\text{GHO}}_{i,t}$	0.81	0.14	(0.52	1.09)	0.51	33
$\Delta \sum_{i=1}^{51} c^{\text{ZHOU}}_{i,t}$	0.79	0.13	(0.54	1.05)	0.55	33
$\Delta \sum_{i=1}^{51} c^{\text{ZHOUG}}_{i,t}$	0.76	0.08	(0.59	0.93)	0.72	33

Table 3.3: Summary statistics for all datasets between 1978 and 1996

Variable	Obs	Mean	Std. Dev.	Min	Max
$\Delta c^{\text{HS}}_{i,t}$	342	0.003	0.045	-0.159	0.123
$\Delta c^{\text{SMM}}_{i,t}$	342	0.003	0.043	-0.126	0.122
$\Delta c^{\text{GHO}}_{i,t}$	342	0.003	0.078	-0.264	0.380
$\Delta c^{\text{ZHOU}}_{i,t}$	342	0.002	0.075	-0.264	0.380
$\Delta c^{\text{ZHOUG}}_{i,t}$	342	0.007	0.059	-0.188	0.192
$\Delta c^{\text{ZHOUG*}}_{i,t}$	318	0.006	0.061	-0.188	0.192

$\Delta c^{\text{ZHOUG}}_{i,t}$: Outliers are set as average growth by state.
$\Delta c^{\text{ZHOUG*}}_{i,t}$: Outliers are set as missing.

Table 3.4: The correlation between $\Delta c_{i,t}^{\mathrm{HS}}$ versus $\Delta c_{i,t}^{\mathrm{SMM}}$, $\Delta c_{i,t}^{\mathrm{GHO}}$ and $\Delta c_{i,t}^{\mathrm{ZHOU}}$, between 1978 and 1996

State	$\rho(\Delta c^{\mathrm{HS}}, \Delta c^{\mathrm{SMM}})$	$\rho(\Delta c^{\mathrm{HS}}, \Delta c^{\mathrm{GHO}})$	$\rho(\Delta c^{\mathrm{HS}}, \Delta c^{\mathrm{Zhou}})$	$\rho(\Delta c^{\mathrm{HS}}, \Delta c^{\mathrm{Zhoug}})$
CA	0.820	0.557	0.557	0.212
FL	0.587	0.685	0.779	0.779
IL	0.770	0.323	0.323	0.592
IN	0.833	0.264	0.264	0.459
LA	0.687	0.325	0.325	0.461
MA	0.747	0.678	0.678	0.678
MD	0.628	0.637	0.637	0.637
MI	0.820	0.275	0.275	0.740
MN	0.752	0.495	0.495	0.419
MO	0.577	0.389	0.389	0.605
NC	0.807	0.442	0.442	0.684
NJ	0.917	0.724	0.724	0.724
NY	0.828	0.695	0.695	0.548
OH	0.962	0.642	0.642	0.297
PA	0.838	0.848	0.848	0.848
TN	0.759	0.618	0.618	0.624
TX	0.779	-0.217	0.137	0.137
VA	0.594	-0.010	0.864	0.864
WI	0.729	0.172	0.172	0.252

Table 3.5: $\Delta c_{i,t}^{*} = \alpha_i + \beta \Delta c_{i,t}^{\mathrm{HS}} + \varepsilon_{i,t}$

$\Delta c_{i,t}^{*}$	β	Std. Error	[95% Conf.	Interval]	$\bar{R}^2$	Obs
$\Delta c_{i,t}^{\mathrm{SMM}}$	0.727	0.036	0.656	0.798	0.531	342
$\Delta c_{i,t}^{\mathrm{GHO}}$	0.732	0.088	0.559	0.906	0.128	342
$\Delta c_{i,t}^{\mathrm{ZHOU}}$	0.756	0.084	0.591	0.920	0.155	342
$\Delta c_{i,t}^{\mathrm{ZHOUG}}$	0.690	0.063	0.567	0.814	0.230	342
$\Delta c_{i,t}^{\mathrm{ZHOUG*}}$	0.771	0.067	0.639	0.904	0.259	318

Table 3.6: $\Delta c_{t}^{*} = \alpha + \beta \Delta c_{t}^{\mathrm{HS}} + \varepsilon_{t}$ for Virginia

Δc_{t}^{*}	β	Std. Error	[95% Conf.	Interval]	$\bar{R}^2$	Obs
$\Delta c_{t}^{\mathrm{SMM}}$	0.732***	0.248	0.207	1.256	0.313	18
$\Delta c_{t}^{\mathrm{GHO}}$	-.015	0.367	-.792	0.763	-.062	18
$\Delta c_{t}^{\mathrm{ZHOU}}$	1.003***	0.146	0.694	1.312	0.732	18

APPENDIX: The attempted reverse-engineering of $C_{i,t}^{\mathbf{SMM}}$

This section examines how national retail sales, likewise state employment and wages, can help predict the annualized growth of C^{SMM}. The aim is to investigate whether these three variables play important roles in the construction of ΔC^{SMM}. If so, it would provide evidence for our claims regarding the construction of C^{SMM}. For comparison purposes, the same experiments are applied to C^{CQS}, C^{GHO}, C^{ZHOU}, and C^{ZHOUG}.

$E_{i,t}$ and $W_{i,t}$ are defined as the level of employment and wages respectively for state i at time t. We then denote the total retail sales data as reported by the quinquennial Census of Retail Trade by C_{i,t_0}^{RCEN}. Furthermore, it is assumed that the Census of Retail Trade at $t = t_0$ is used in the constructions of $C_{i,t}^{\text{SMM}}$, where $t = t_0 + 3$, $t_0 + 4$, ..., $t_0 + 7$. The average annualized growth rate of each variable is then calculated as follows:

$$\Delta\widetilde{C}_{i,t}^{\text{SMM}} = \frac{ln(C_{i,t}^{\text{SMM}}) - ln(C_{i,t_0}^{RCEN})}{t - t_0} \tag{3.1}$$

$$\Delta\widetilde{E}_{i,t} = \frac{ln(E_{i,t}) - ln(E_{i,t_0})}{t - t_0} \tag{3.2}$$

$$\Delta\widetilde{W}_{i,t} = \frac{ln(W_{i,t}) - ln(W_{i,t_0})}{t - t_0} \tag{3.3}$$

$$\Delta\widetilde{retail}_t^{\text{U.S.}} = \frac{ln(retail_t^{\text{U.S.}}) - ln(C_{i,t_0}^{RCEN})}{t - t_0} \tag{3.4}$$

The following estimation equation is tested:

$$\Delta\widetilde{C}_{i,t}^{\text{SMM}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t} \text{ for } \Delta\widetilde{C}_{i,t}^{\text{SMM}}, \text{ which utilizes } C_{i,t_0}^{RCEN} \tag{3.5}$$

As a starting point, this paper focuses on the 19 states where $\Delta\widetilde{C}_{i,t}^{\text{HS}}$ is available and makes comparisons across the growth rates of all of the different measures of retail sales. The regression results for equation (3.5) are presented by census year. Again, assuming C^{HS} as the benchmark, Table 3.7 shows that a moderate proportion of the variations in $\Delta\widetilde{C}_{i,t}^{\text{HS}}$ could be explained by wage, employment and aggregate retail sales growth. However, $\Delta\widetilde{C}_{i,t}^{\text{SMM}}$ shows much higher dependencies on the three variables for two out of the four periods, and $\Delta\widetilde{C}_{i,t}^{\mathrm{CQS}}$ even shows R-Squares above 90 percent for some periods. These results suggest that employment and wages contribute a great deal to the construction of C^{SMM} and C^{CQS}. By contrast, Tables 3.10, 3.11 and 3.12 show that a larger proportion of the variations in C^{GHO}, C^{ZHOU} and C^{ZHOUG} can not be explained by these three variables.

We then extend our examination to all states and periods where applicable. Since Claritas claims that local sales tax data are also employed in its construction of the annualized growth of retail sales, whenever available, $C_{i,t}^{\mathrm{GHO}}$ is added to the estimation equations for $C_{i,t}^{\mathrm{SMM}}$ and $C_{i,t}^{\mathrm{CQS}}$. Consequently, two additional equations are tested:

$$\Delta\widetilde{C}_{i,t}^{\mathrm{SMM}} = \alpha + \beta_1\Delta\widetilde{C}_{i,t}^{\mathrm{GHO}} + \beta_2\Delta\widetilde{E}_{i,t} + \beta_3\Delta\widetilde{W}_{i,t} + \beta_4\Delta\widetilde{retail}_t^{US}, \text{ if } C_{i,t}^{\mathrm{GHO}} \text{ is not missing} \tag{3.6}$$

$$\Delta\widetilde{C}_{i,t}^{\mathrm{SMM}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{US}, \text{ if } C_{i,t}^{\mathrm{GHO}} \text{ is missing} \tag{3.7}$$

As shown before, these three variables continue to explain a larger proportion of variations in $\Delta\widetilde{C}_{i,t}^{\mathrm{SMM}}$ and $\Delta\widetilde{C}_{i,t}^{\mathrm{CQS}}$ than $\Delta\widetilde{C}_{i,t}^{\mathrm{GHO}}$, $\Delta\widetilde{C}_{i,t}^{\mathrm{ZHOU}}$, and $\Delta\widetilde{C}_{i,t}^{\mathrm{ZHOUG}}$. Furthermore, $\Delta\widetilde{C}_{i,t}^{\mathrm{CQS}}$ stands out again, as roughly 80 percent of its variance can be sufficiently explained by this simple linear regression, for three out of four periods.

Another interesting observation is that $\Delta\widetilde{C}_{i,t}^{\mathrm{SMM}}$ and $\Delta\widetilde{C}_{i,t}^{\mathrm{CQS}}$ tend to be better explained in states where $\Delta\widetilde{C}_{i,t}^{\mathrm{GHO}}$ is not available. This observation is reflected by the higher R-Squares in Tables 3.15 and 3.18 relative to those in Tables 3.14 and 3.17. This seems counter-intuitive at first sight, as $\Delta\widetilde{C}_{i,t}^{\mathrm{GHO}}$ is expected to be positively correlated with $\Delta\widetilde{C}_{i,t}^{\mathrm{SMM}}$ and $\Delta\widetilde{C}_{i,t}^{\mathrm{CQS}}$ and to provide state-specific retail sales-related information beyond employment, wage and aggregate retail growth. Actually, both Table 3.14 and Table 3.17 present significantly positive correlations between $\Delta\widetilde{C}_{i,t}^{\mathrm{GHO}}$, $\Delta\widetilde{C}_{i,t}^{\mathrm{SMM}}$ and $\Delta\widetilde{C}_{i,t}^{\mathrm{CQS}}$. The lower R-Squares for most periods, however, might be justified by the fact that, given that there is less state-specific retail sales-related information, the constructions of $\Delta\widetilde{C}_{i,t}^{\mathrm{SMM}}$ and $\Delta\widetilde{C}_{i,t}^{\mathrm{CQS}}$ are more constrained by data limitations, and therefore are more heavily dependent on employment, wages, and aggregate retail growth.

Another fact worth noting is that the coefficient of each variable varies across census years. Readers should keep in mind that the exact formula used in the estimation of $\Delta C^{\text{SMM}}_{i,t}$ is not revealed to the public, and might adopt a very complicated nonlinear form. This study has no intention of claiming that Equations (3.5), (3.6), and (3.7) are the exact ones employed by Claritas in constructing $\Delta C^{\text{SMM}}_{i,t}$. The purpose of this experiment is to find supporting evidence for the information the author inferred from the limited available sources.[17]

Table 3.7: $\Delta \widetilde{C}^{\text{HS}}_{i,t} = \alpha + \beta_1 \Delta \widetilde{E}_{i,t} + \beta_2 \Delta \widetilde{W}_{i,t} + \beta_3 \Delta \widetilde{retail}^{\text{U.S.}}_{t} + \varepsilon_{i,t}$

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta \widetilde{E}_{i,t}$	0.25	0.61*	1.18***	1.48**
	(0.35)	(0.36)	(0.3)	(0.68)
$\Delta \widetilde{W}_{i,t}$	0.85***	0.17	0.33	-.03
	(0.24)	(0.21)	(0.33)	(0.58)
$\Delta \widetilde{retail}_{\text{U.S.},t}$	-.06	1.74***	0.11	4.21
	(0.29)	(0.35)	(0.21)	(4.09)
Obs.	95	95	95	38
$\bar{R}^2$	0.65	0.43	0.45	0.4

Table 3.8: $\Delta \widetilde{C}^{\text{SMM}}_{i,t} = \alpha + \beta_1 \Delta \widetilde{E}_{i,t} + \beta_2 \Delta \widetilde{W}_{i,t} + \beta_3 \Delta \widetilde{retail}^{US}_{t} + \varepsilon_{i,t}$, where $\Delta \widetilde{C}^{\text{HS}}_{i,t}$ is present

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta \widetilde{E}_{i,t}$	0.7***	-.44**	0.93***	1.51**
	(0.21)	(0.19)	(0.23)	(0.68)
$\Delta \widetilde{W}_{i,t}$	0.29**	0.83***	0.16	-.03
	(0.14)	(0.11)	(0.25)	(0.57)
$\Delta \widetilde{retail}_{\text{U.S.},t}$	0.34**	1.96***	-.11	6.35
	(0.17)	(0.18)	(0.16)	(4.04)
Obs.	95	95	95	38
$\bar{R}^2$	0.77	0.79	0.39	0.43

Standard errors in parenthesis. {*, **, ***} = significant at the {10%, 5%, 1%} level.

[17] Several other regression forms, such as those with interaction terms and/or nonlinear terms, were also studied. The results are similar, and are thus not reported.

Table 3.9: $\Delta\widetilde{C}_{i,t}^{\text{CQS}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$, where $\Delta\widetilde{C}_{i,t}^{\text{HS}}$ is present

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta\widetilde{E}_{i,t}$	0.52*** (0.19)	0.18 (0.16)	1.44*** (0.14)	1.22** (0.53)
$\Delta\widetilde{W}_{i,t}$	0.51*** (0.13)	0.71*** (0.09)	-.02 (0.15)	-.02 (0.45)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	0.15 (0.45)	0.85*** (0.16)	-.09 (0.09)	2.34 (3.15)
Obs.	57	95	95	38
$\bar{R}^2$	0.94	0.85	0.78	0.42

Table 3.10: $\Delta\widetilde{C}_{i,t}^{\text{GHO}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$, where $\Delta\widetilde{C}_{i,t}^{\text{HS}}$ is present

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta\widetilde{E}_{i,t}$	0.88 (0.54)	-2.79*** (0.74)	-2.50*** (0.84)	-1.26 (3.05)
$\Delta\widetilde{W}_{i,t}$	0.83** (0.37)	2.79*** (0.43)	1.54* (0.91)	-.83 (2.59)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-.56 (0.45)	1.02 (0.73)	0.83 (0.57)	-2.52 (18.25)
Obs.	95	95	95	38
$\bar{R}^2$	0.59	0.46	0.08	-.003

Table 3.11: $\Delta\widetilde{C}_{i,t}^{\text{ZHOU}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$, where $\Delta\widetilde{C}_{i,t}^{\text{HS}}$ is present

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta\widetilde{E}_{i,t}$	0.92* (0.54)	-2.80*** (0.74)	-2.67*** (0.84)	-1.34 (3.16)
$\Delta\widetilde{W}_{i,t}$	0.82** (0.37)	2.77*** (0.43)	1.59* (0.91)	-.64 (2.68)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-.61 (0.45)	0.91 (0.73)	0.97* (0.57)	-1.62 (18.88)
Obs.	95	95	95	38
$\bar{R}^2$	0.6	0.45	0.1	-.02

Standard errors in parenthesis. {*, **, ***} = significant at the {10%, 5%, 1%} level.

Table 3.12: $\Delta\widetilde{C}_{i,t}^{\text{ZHOUG}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$, where $\Delta\widetilde{C}_{i,t}^{\text{HS}}$ is present

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta\widetilde{E}_{i,t}$	0.59 (0.51)	-1.66** (0.8)	0.11 (0.8)	-2.16 (2.83)
$\Delta\widetilde{W}_{i,t}$	0.31 (0.35)	1.76*** (0.46)	-.49 (0.87)	1.25 (2.39)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	0.53 (0.42)	2.45*** (0.79)	1.23** (0.55)	15.67 (16.90)
Obs.	95	95	95	38
$\bar{R}^2$	0.33	0.29	0.02	-.05

Table 3.13: $\Delta\widetilde{C}_{i,t}^{\text{SMM}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$

	t_0=1972	t_0=1977	t_0=1982	t_0=1987	t_0=1992	t_0=1997
$\Delta\widetilde{E}_{i,t}$	0.19 (0.13)	-.007 (0.12)	-.26** (0.1)	1.10*** (0.14)	0.96*** (0.15)	0.37 (0.24)
$\Delta\widetilde{W}_{i,t}$	0.41*** (0.08)	0.58*** (0.09)	0.72*** (0.06)	0.07 (0.13)	0.18 (0.12)	0.58*** (0.16)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-.49** (0.21)	0.34** (0.15)	2.03*** (0.16)	-.08 (0.11)	2.00*** (0.31)	1.65*** (0.33)
Obs.	255	250	255	255	255	255
$\bar{R}^2$	0.47	0.54	0.7	0.52	0.58	0.38

Table 3.14: $\Delta\widetilde{C}_{i,t}^{\text{SMM}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t}\beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \beta_4\Delta\widetilde{C}_{i,t}^{\text{GHO}} + \varepsilon_{i,t}$, where $\Delta\widetilde{C}_{i,t}^{\text{GHO}}$ is present

	t_0=1972	t_0=1977	t_0=1982	t_0=1987	t_0=1992	t_0=1997
$\Delta\widetilde{E}_{i,t}$	0.32** (0.15)	-.37*** (0.14)	-.52*** (0.13)	1.12*** (0.16)	0.75*** (0.18)	0.57** (0.29)
$\Delta\widetilde{W}_{i,t}$	0.46*** (0.11)	0.65*** (0.11)	0.88*** (0.09)	0.14 (0.16)	0.43*** (0.15)	0.47** (0.18)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-.77*** (0.21)	0.34** (0.15)	2.02*** (0.17)	-.11 (0.12)	2.26*** (0.33)	1.64*** (0.35)
$\Delta\widetilde{C}_{i,t}^{GHO}$	-.05 (0.04)	0.13*** (0.03)	-.03 (0.02)	0.003 (0.02)	-.03 (0.02)	-.03 (0.02)
Obs.	225	220	225	225	225	225
$\bar{R}^2$	0.42	0.62	0.7	0.48	0.58	0.36

Standard errors in parenthesis. {*, **, ***} = significant at the {10%, 5%, 1%} level.

Table 3.15: $\Delta\widetilde{C}^{\text{SMM}}_{i,t} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}^{\text{U.S.}}_t + \varepsilon_{i,t}$, where $\Delta\widetilde{C}^{\text{GHO}}_{i,t}$ is missing

	t_0=1972	t_0=1977	t_0=1982	t_0=1987	t_0=1992	t_0=1997
$\Delta\widetilde{E}_{i,t}$	0.11	0.78***	0.57	1.18***	1.58***	-.41
	(0.29)	(0.26)	(0.4)	(0.43)	(0.29)	(0.56)
$\Delta\widetilde{W}_{i,t}$	0.46***	0.01	0.49***	-.22	-.49**	1.68***
	(0.17)	(0.18)	(0.18)	(0.37)	(0.23)	(0.48)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	0.85	0.19	2.21***	0.08	1.04	1.36
	(0.8)	(0.42)	(0.56)	(0.28)	(0.78)	(0.99)
Obs.	30	30	30	30	30	30
$\bar{R}^2$	0.64	0.53	0.76	0.64	0.69	0.59

Table 3.16: $\Delta\widetilde{C}^{\text{CQS}}_{i,t} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}^{\text{U.S.}}_t + \varepsilon_{i,t}$

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta\widetilde{E}_{i,t}$	0.76***	0.25***	1.83***	1.17***
	(0.13)	(0.1)	(0.1)	(0.13)
$\Delta\widetilde{W}_{i,t}$	0.27***	0.81***	-.33***	-.27**
	(0.1)	(0.05)	(0.09)	(0.11)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-.58	0.74***	-.05	1.37***
	(0.45)	(0.14)	(0.07)	(0.28)
Obs.	147	253	255	255
$\bar{R}^2$	0.79	0.84	0.79	0.54

Table 3.17: $\Delta\widetilde{C}^{\text{CQS}}_{i,t} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}^{\text{U.S.}}_t + \beta_4\Delta\widetilde{C}^{\text{GHO}}_{i,t} + \varepsilon_{i,t}$, where $\Delta\widetilde{C}^{\text{GHO}}_{i,t}$ is present

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta\widetilde{E}_{i,t}$	0.32*	0.18	1.90***	1.25***
	(0.18)	(0.13)	(0.11)	(0.17)
$\Delta\widetilde{W}_{i,t}$	0.58***	0.91***	-.40***	-.38***
	(0.13)	(0.08)	(0.11)	(0.15)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	0.22	0.84***	-.04	1.22***
	(0.5)	(0.15)	(0.08)	(0.32)
$\Delta\widetilde{C}^{GHO}_{i,t}$	0.008	-.07***	0.009	0.02
	(0.03)	(0.02)	(0.01)	(0.01)
Obs.	129	223	225	225
$\bar{R}^2$	0.8	0.84	0.75	0.47

Standard errors in parenthesis. {*, **, ***} = significant at the {10%, 5%, 1%} level.

Table 3.18: $\Delta\widetilde{C}_{i,t}^{\text{CQS}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$, where $\Delta\widetilde{C}_{i,t}^{\text{GHO}}$ is missing

	t_0=1977	t_0=1982	t_0=1987	t_0=1992
$\Delta\widetilde{E}_{i,t}$	1.43***	0.08	1.45***	1.41***
	(0.31)	(0.34)	(0.27)	(0.19)
$\Delta\widetilde{W}_{i,t}$	-.27	0.99***	0.009	-.26*
	(0.26)	(0.16)	(0.24)	(0.14)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-2.25	0.77	-.17	1.64***
	(1.53)	(0.49)	(0.18)	(0.49)
Obs.	18	30	30	30
$\bar{R}^2$	0.75	0.87	0.91	0.88

Table 3.19: $\Delta\widetilde{C}_{i,t}^{\text{GHO}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$

	t_0=1972	t_0=1977	t_0=1982	t_0=1987	t_0=1992	t_0=1997
$\Delta\widetilde{E}_{i,t}$	0.29	0.02	-1.43***	-2.53***	-3.20***	-.89
	(0.25)	(0.35)	(0.42)	(0.51)	(0.62)	(1.12)
$\Delta\widetilde{W}_{i,t}$	1.13***	1.37***	2.04***	3.25***	2.52***	2.38***
	(0.16)	(0.25)	(0.25)	(0.48)	(0.51)	(0.71)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-.43	-.51	1.41***	0.17	2.09	-1.70
	(0.34)	(0.36)	(0.53)	(0.39)	(1.31)	(1.36)
Obs.	225	220	225	225	270	225
$\bar{R}^2$	0.57	0.51	0.46	0.17	0.08	0.11

Table 3.20: $\Delta\widetilde{C}_{i,t}^{\text{ZHOU}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$

	t_0=1972	t_0=1977	t_0=1982	t_0=1987	t_0=1992	t_0=1997
$\Delta\widetilde{E}_{i,t}$	0.29	0.04	-1.44***	-2.59***	-3.14***	-1.14
	(0.25)	(0.35)	(0.42)	(0.51)	(0.66)	(1.12)
$\Delta\widetilde{W}_{i,t}$	1.13***	1.37***	2.03***	3.28***	2.51***	2.65***
	(0.16)	(0.25)	(0.25)	(0.48)	(0.55)	(0.7)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-.43	-.54	1.36***	0.21	2.18	-1.97
	(0.34)	(0.36)	(0.53)	(0.39)	(1.41)	(1.36)
Obs.	225	220	225	225	270	225
$\bar{R}^2$	0.57	0.52	0.45	0.17	0.07	0.13

Standard errors in parenthesis. {*, **, ***} = significant at the {10%, 5%, 1%} level.

Table 3.21: $\Delta\widetilde{C}_{i,t}^{\text{ZHOUG}} = \alpha + \beta_1\Delta\widetilde{E}_{i,t} + \beta_2\Delta\widetilde{W}_{i,t} + \beta_3\Delta\widetilde{retail}_t^{\text{U.S.}} + \varepsilon_{i,t}$

	t_0=1972	t_0=1977	t_0=1982	t_0=1987	t_0=1992	t_0=1997
$\Delta\widetilde{E}_{i,t}$	0.72*** (0.22)	0.19 (0.35)	-.68 (0.5)	-.33 (0.54)	-.83 (0.66)	-2.43** (1.07)
$\Delta\widetilde{W}_{i,t}$	0.63*** (0.15)	1.02*** (0.25)	1.30*** (0.3)	0.99* (0.51)	0.55 (0.55)	2.86*** (0.67)
$\Delta\widetilde{retail}_{\text{U.S.},t}$	-.58* (0.31)	0.08 (0.37)	2.45*** (0.63)	1.14*** (0.41)	4.13*** (1.41)	-.28 (1.30)
Obs.	225	220	225	225	270	225
$\bar{R}^2$	0.52	0.43	0.27	0.06	0.02	0.1

Standard errors in parenthesis. {*, **, ***} = significant at the {10%, 5%, 1%} level.

Chapter 4

Measuring Wealth Effects Using U.S. State Data

4.1 Introduction

During the second half of the 1990s, a skyrocketing stock market boosted the wealth holdings of American households; at the same time, the personal saving rate dropped from about 8 to 2 percent. This so-called "saving rate puzzle" sparked renewed policy and research interest in the wealth effects on consumption. Figure 4.1 shows a relatively stylized negative correlation between the saving rate and the net worth to income ratio, which implies a positive correlation between wealth and consumption after controlling for the income effect.

If it is the rise in wealth that is driving down the personal saving rate, we should expect that future variations in wealth will have an impact on consumption. Consequently, wealth effects should be taken into consideration when implementing monetary policy. We should be skeptical, however, about the seemingly obvious relationship between consumption and wealth for a variety of reasons. First, the association we observe in Figure 4.1 could be mainly the result of simultaneity. For instance, any shock to consumers' optimism or pessimism could have an impact on housing prices, stock prices, and consumption growth in the same direction. Second, endogeneity could also be triggered by a reverse causality of consumption on wealth. Given the presence of heterogeneity, aggregation is another problem, as summing up individuals might not produce a representative consumer. In addition, measurement errors could lead to unreliable associations. To give an example, assume that income Y is measured with error. Through construction then, the personal saving rate, $s = 1 - C/Y$, will also be mis-measured in the same direction as Y. At the same time, the measured wealth-income ratio W/Y will be biased in the opposite direction. The measurement error in income will thus induce a negative correlation between the saving rate and the worth to income ratio.

Most of the current literature on wealth effects employs either aggregate or household-level data. Studies using aggregate data are subject to endogeneity and aggregation problems. On the other hand, studies using household-level data suffer from serious measurement error problems. There is, in fact, a very limited choice of household-level data available for carrying out such studies. For instance, the *Panel Study on Income Dynamics* (PSID) only measures food consumption, while the *Consumer Expenditure Survey* (CEX) has detailed but noisy data on household expenditures and poor financial information. *The Survey of Consumer Finances* (SCF) provides no measure of consumption at all.

An alternative approach, one that potentially avoids some of the problems related to both aggregate and household-level data, is to utilize regional variation. First, aggregation is likely to be less of a problem when less aggregated data is used. Second, if there is sufficient variation across regions, the endogeneity problem might be better controlled. For instance, let us assume a region-specific shock to consumers' confidence, one that might also have a large impact on the consumption behavior of households in the region. However, if a well-integrated stock market exists, this region-specific shock might not have as great an impact on regional stock prices as an aggregate shock would. Therefore, the endogeneity problem is alleviated to some extent. On the other hand, it can be argued that regional data provides more comprehensive and better measures of the relevant variables than household-level data. Furthermore, regional data is more likely to cover a longer time period and therefore allow for richer dynamics.

Case, Quigley, and Shiller (2005) did pioneering work using U.S. state-level data to estimate and compare housing wealth effects and stock wealth effects. This paper extends their work in several aspects. We construct a new panel dataset of financial wealth for U.S. states, using anonymous proprietary account-level records of geographic wealth holdings. The new dataset is more comprehensive and representative than existing alternative measures. This paper also improves upon Case, Quigley, and Shiller (2005), in that we construct a significantly improved state-level proxy for consumption data. These datasets are then combined to provide new estimates of the wealth effects on consumption from changes in stock and housing wealth. The rest of the paper is organized as follows: Section 2 reviews the related literature; Section 3 discusses the limitations of the currently available state-level consumption and stock wealth datasets; Section 4 describes the newly constructed data; Section 5 presents the model specification and regression results; and Section 6 concludes.

4.2 Recent evidence

The current literature on the marginal propensity to consume (MPC) out of different components of wealth is limited. Davis and Palumbo (2001) compared the stock market wealth effect with the non-stock market wealth effect using U.S. aggregate data. The results, derived from a co-integration analysis, are, however, sensitive to model specifications. Specifically, the long-run effects of both types of wealth are about the same (i.e., 0.06 for stocks and 0.08 for non-stocks) when the level of variables is used. Using logarithms, however, the results show an elasticity for non-stock wealth four times greater than that for stock wealth; this implies that the MPC out of non-stock wealth is at least twice as large as the MPC out of stock wealth. Additionally, using aggregate data (though applying a different method), Carroll, Otsuka, and Slacalek (2006) reported an immediate MPC out of housing wealth of about 1.5 cents and an immediate stock wealth MPC of 0.75. The difference, however, is found to be statistically insignificant from zero.

Levin (1998) appears to be the first study in the U.S. that using household-level data to estimate the differential effect of housing and stock wealth. Using the Retirement History Survey, Levin found that housing wealth has essentially no effect on consumption. Out of eight spending categories, only three reported a statistically significant difference between the respective coefficients for liquid and housing wealth. This finding contradicts the studies using aggregate data summarized above. A possible reason could be the fact that every interviewee in the survey is at least 65 years old. If elderly people tend to view housing wealth more as consumption than as an investment item, their housing wealth effect will be lower than would otherwise be the case. Using the CEX and SCF, Bostic, Gabriel, and Painter (2005) find that, while incorporating all households in their sample, there is no evidence for an important housing wealth effect. Among home owners, however, the housing wealth elasticity is found to be consistently significant and larger than the stock wealth elasticity. Their paper also suggests different consumption behaviors for credit-constrained versus non credit-constrained samples.

Among those who use panel data, Case, Quigley, and Shiller (2005) are probably the most cited in the current literature. Using quarterly U.S. state-level data for 1982 through 1999, the authors found a significant housing wealth elasticity of about 5 percent, but an economically negligible stock wealth elasticity under most model specifications. When using a panel of annual data for 14 developed countries, they found an even larger housing wealth elasticity, in the range of $11 - 15$ percent. Nonetheless, under all cases, they found no evidence for an important stock wealth effect. Bayoumi and Edison (2003) used data for 16 industrial countries and found significant wealth effects for most samples and periods. Their estimated housing wealth effect was consistently larger than their estimated equity wealth effect. Ludwig and Sløk (2002) found evidence contrary to the studies cited above. Using annual data from 16 OECD countries, and taking housing prices and stock market prices as proxies for their respective wealth components, the authors reported an estimated stock wealth elasticity twice the estimated housing wealth elasticity. Additionally, both estimates were found to be positive and statistically significant. On the other hand, Girouard and Blöndal (2001) also used OECD data, but were unable to arrive at consistent results when comparing housing wealth with financial wealth. Dvornak and Kohler (2003), using Australian state-level data, found a larger stock wealth effect than housing wealth effect.

4.3 Limitations of existing state-level consumption and stock wealth data

Case, Quigley, and Shiller (2005) have constructed the only measure of quarterly state-level stock wealth for the U.S. for the period 1982 through 1999. They obtained annual information on mutual fund holdings at the state level, which is only available for the years 1986, 1987, 1989, 1991, and 1993. In order to construct stock wealth data, the authors needed to make two very restrictive assumptions. First, they assumed that the proportion of mutual funds out of financial assets was constant. However, Figure 4.2 plots the proportion of mutual funds out of total stock wealth, and shows an evident increase in that proportion over time. Second, they assumed a constant asset distribution across states for those years during which mutual fund data were not available. During those years, then, the stock wealth of each state should, based on the construction, mimic the movement of aggregate stock wealth. Given the absence of any real wealth distribution across states, the dataset they created is not "state-level" data.

To the best of my knowledge, there exist three distinctive state-level consumption datasets – those used by Asdrubali, Sorensen, and Yosha (1996); Case, Quigley, and Shiller (2005); and Garrett, Hernàndez-Murillo, and Owyang (2004). Of these, only Case, Quigley, and Shiller (2005) utilized the data to examine wealth effects. The consumption data used by Asdrubali, Sorensen, and Yosha (1996), and Case, Quigley, and Shiller (2005) were constructed from retail sales based on different private sector sources. However, in both cases, the quality of the data derived from the private sources is questionable, for a variety of reasons. First, the methodology used in the data construction is never explicitly revealed by either private source. Second, retail sales are presented for states that do not implement sales tax, which constitutes perhaps the single most important source for calculating state retail sales after the Census Bureau ceased reporting monthly retail sales by state, in 1997.

Last but not least, both sources vaguely note that important state variables like wage and employment are incorporated into the estimation of retail sales. As a result, the datasets will induce unreliable estimations of the relationship between consumption and any variable that is correlated with wage or employment.

Garrett, Hernàndez-Murillo, and Owyang (2004) computed quarterly retail sales by dividing sales tax revenue by the sales tax rate. The data is potentially a good measure of state retail sales, and thus is generally adopted in this paper. One problem with this, however, is that the sales tax revenues are measured with serious errors; this results in unreasonably large consumption variations and apparent outliers. Therefore, this paper improves upon the data used in Garrett et. al. (2004) by constructing more accurate measures of state retail sales, and by explicitly accounting for outliers.

4.4 Data description

This paper uses a panel dataset for 44 U.S. states as well as Washington, D.C., at a semi-annual frequency for the period 2001 through 2005. The newly constructed datasets are for stock wealth and consumption at the state level. Other important variables include after-tax labor income and housing wealth. All are expressed in real per capita terms. There is evidence that the new data is more comprehensive and accurate than other existing alternatives. Some important findings will be discussed in the rest of this section. More detailed discussions can be found in the second and third chapter of this dissertation.

4.4.1 Stock wealth data

The author obtained anonymous account-level records on financial wealth holdings at the ZIP+4 Code level from the IXI Corporation. At the end of each semiannual cycle, IXI collects data from more than 85 leading financial institutions in its network, IXI►Net™. Reporting institutions include major banks, brokerage firms, insurance companies and mutual fund dealers. Additional information can be found in Chapter 2 of this dissertation.

Stock market wealth is defined as the sum of directly and indirectly held (i.e., investments, in the form of IRA and Keogh accounts) stocks and mutual funds. Stock wealth growth is constructed using a consistent method for all 50 states plus the District of Columbia.[1] The geographic distribution of stock wealth growth is plotted in Figure 2.3. We find similar patterns across states, something to be expected given the fact that the U.S. stock market is so well integrated. However, whether the state heterogeneity manifested in the figure reflects reality cannot be readily answered, as there exists no alternative state-level wealth data with which we might make comparison.

Nevertheless, there are some stylized facts about the U.S. that could help us make a judgment. Florida and Arizona are the two states that have the highest percentage of retired people. As reflected in Figure 2.3, their seasonal patterns also distinguish them from other states. In order to better illustrate the differences, Figures 4.3(a) and 4.3(b) compare the stock wealth growth of Florida and Arizona with the average stock wealth growth of the other states. Both figures indicate that Florida and Arizona have a much higher stock wealth growth rate than the other states during the second half of each year, and a much lower stock wealth growth rate during the first half of each year. This phenomenon might seem strange at first glance, but is actually an outcome of the "snow-bird effect." In the U.S., retired people tend to move to Florida and Arizona during the winter and then move back to their permanent residences once the winter is over. If such individuals update their physical mailing addresses with their financial institutions each time they relocate, they effectively bring their assets along with them.[2] Along with a single measure of population over the course of one year, we should expect that the "snow bird" effect to be fully captured by stock wealth growth at semiannual frequencies. Figures 4.3(a) and 4.3(b) therefore provide another piece of evidence that the heterogeneity found in the data corresponds to reality.

A substantial effort was extended to find other potential state-level financial resources with which the new data could be compared. Thus, for instance, Bloomberg reports local stock indices for 22 states, the growth of which is expected to positively but not perfectly

[1] Details on its construction can be found in Chapter 2 of this dissertation.

[2] As per the practice of the IXI corporation, the assets are now considered as belonging to the Zip Code +4 of the updated new address.

correlate with local stock wealth growth. Figure 4.4 presents the correlation between the local stock index and local stock wealth, broken down graphically. Out of the 23 calculated correlations, we find only 2 negative numbers. At the state-specific growth level, defined as state growth minus the U.S. national component, there are still 15 positive correlations. These facts further provide supporting evidence that the data reflects a true distribution of stock market wealth across states.

4.4.2 Consumption data

Since measures of personal consumption expenditure (PCE) at the state level are not available in the U.S., retail sales are used as a proxy for consumption. In the U.S., national retail sales account for roughly half of PCE, and *The Retail Trade Survey* is probably the single most important source for the national PCE estimation carried out by the Bureau of Economic Analysis (BEA).[3] These considerations provide us with a rationale for using retail sales in place of consumption.

However, even retail sales data is not directly available in the U.S. at the state level. Following Garrett, Hernàndez-Murillo, and Owyang (2004), quarterly state-level general sales tax revenues can be obtained from the *Quarterly Summary of State and Local Government Tax Revenue*, published by the U.S. Census Bureau. Together with general sales tax rates collected from various sources,[4] state-level retail sales are computed by dividing the state general sales tax revenue by the general sales tax rate. One limitation of this method is that it can be applied to 45 states and the District of Columbia. Nevada, however, is dropped in this study because of its discontinued data report and obvious poor data quality.

Strictly speaking, the computed retail sales are only one component of real retail sales, as they exclude items that are either not subject to sales tax or are part of special tax programs, i.e., liquor and cigarettes. Furthermore, there is serious measurement error problem with the computed retail sales. The author, however, found state-level government-reported

[3] See Wilcox (1992).

[4] The state general sales tax rate can be found from various sources such as the *State Government Tax Collections*, and the Tax Foundation's *Facts and Figures on Government Finances*.

(taxable) retail sales for 12 states for the same period during which state-level stock wealth data is available.[5] These measures are more comprehensive than the computed retail sales, as they either include all consumption items (such as when government-reported gross retail sales are used) or at least include those items that are part of special tax programs.[6] Furthermore, these government-reported measures should be more accurate and reliable than the computed ones, since local governments have access to more information regarding their own sales tax system and tax collection practices than other people do.[7]

Ideally, government-reported (taxable) retail sales should be used as a measure of consumption. However, since they are only available for a limited number of states, this paper compiles three sets of consumption data according to the quality of the retail sales data. The first one includes those 12 states that have government-reported retail sales or taxable retail sales; it is categorized as "Best Data". The second set is called "Combined Data,"[8] and includes "Best Data" along with the computed retail sales for the other states. The third set is called "Good Data," which includes "Combined Data" with outliers taken care of. Please refer to the third chapter of this dissertation for a more detailed discussion of the consumption data.

4.4.3 Data from other sources

Other important variables used in this paper include quarterly after-tax labor income and housing wealth. After-tax labor income is calculated following Lettau and Ludvigson (2001). The formula used to construct state-level housing wealth is similar to the one adopted by Case, Quigley, and Shiller (2005), and is given as follows:

$$w^h_{i,t} = (HO_{i,t} * HH_{i,t}) * HPI_{i,t} * HV_i,$$

[5] Data are obtained from the websites of the respective state tax administrations.

[6] Special tax programs notably constitute roughly 25 percent of total sales tax revenue.

[7] They are either calculated by local governments (as in Virginia), or are derived directly from the reports on dealers' returns (as in Iowa).

[8] This paper also examined the wealth effects using another set of dataset that only incorporates the computed taxable retail sales. Please refer to Table 4.3 for discussions of the results.

where w_i^h is the value of the owner occupied housing wealth for state i; HO is the home ownership rate, taken from the Census Bureau; HPI is the weighted repeat sales housing price index, taken from the Federal Housing Finance Agency (FHFA); and HV is the average home price for 1999, taken from the 2000 Census.

4.4.4 Data issues

One important data issue arises here. As mentioned above, all variables except the stock market wealth are available at quarterly frequencies. To make them analogous to the stock market wealth, this paper takes their means over the quarters for each half-year, thus converting them into semiannual frequencies.

The dataset, however, features evident and sizable seasonal patterns at the semiannual frequency, especially for the constructed consumption data. The author has made a considerable effort at removing them in a consistent fashion, but was unable to do so at the semiannual frequency. This is largely because of the heterogeneity of seasonal patterns across states and the relatively short time horizon. Nevertheless, many state governments recommend using longer time spans for more reliable trends. It should be recognized that measures of taxable sales (or revenue) at higher frequencies could be misrepresentative for the purpose of comparison. This is because of timing errors over the year-long period. The above consideration recommends using annual growth rates so as to eliminate seasonal effects, at the cost of fewer observations and thus a reduced regression power.

Additionally, to avoid a time aggregation problem, annual averages are not used to calculate growth rates. Instead, $\Delta c_{i,t}$ is computed as the log difference between consumptions for the first half of year t and for year $t-1$. The first half was chosen in consideration of the fact that the state fiscal year ends on June 30. It is arguable that data collected towards the end of a fiscal year is more accurate than data collected at any other time of year.

4.4.5 Another look at the new data

Since this paper relies heavily on the two newly constructed datasets, before examining the wealth effects, the data is again examined closely by estimating the following equation:

$$\Delta c_{i,t} = \alpha_t + \beta_1 \Delta y_{i,t} + \beta_2 \Delta w_{i,t}^f + \beta_3 \Delta w_{i,t}^h + \varepsilon_{i,t}, \tag{4.1}$$

where Δ denotes the growth rate of a variable, i.e., the log difference of the variable in real per capita terms. Equation 4.1 is a simple description of the data without taking into consideration simultaneity and aggregation problems. Table 4.1 reports the results for all three datasets. It shows that income growth is the one variable that consistently has the largest and most significant coefficient. Perhaps the most interesting finding is that there is evidence that consumption positively correlates with the growth rates of both housing wealth and stock wealth when they are regressed separately. Conversely, whenever income growth is included, their respective coefficients become much less significant, in connection with the reduced sizes. The data archive that can produce all results in this study is available from Johns Hopkins library, at URL: http://jhir.library.jhu.edu/handle/1774.2/34267.[9]

4.5 Regressions

4.5.1 Wealth effect estimations

Most studies in the current literature, particularly those that focus on the immediate response of consumption to wealth, adopt regressions similar to those used in Equation 4.1.[10] However, such regressions do not yield straightforward wealth effects, since they only report the contemporaneous percentage correlation between consumption and wealth. Worse, tests

[9] Instructions on how to obtain the new data of financial wealth growth rate for U.S. states can be found in the read me file for the data archive.

[10] Cointegration analysis is another standard method used in the current literature to study long-term MPCs. Nevertheless, given the relatively short time horizon, the data used in this paper does not allow for such an analysis. Additionally, cointegration analysis is intrinsically problematic. The most relevant problem with respect to income and wealth effect analysis is the requirement that the cointegrating vectors remain stable, which in turn requires a stable saving rate. This requirement, however, obviously runs contrary to what the data tells us, as illustrated in Figure 4.1.

of equal stock and housing wealth effects do not produce transparent results.[11] In order to solve this problem, this paper adopts an approach similar to that employed by Carroll, Otsuka, and Slacalek (2006), wherein they use the ratio of the change in each variable relative to an initial level of after-tax labor income. Put another way, if we define

$$\begin{aligned}
\Delta \tilde{c}_{i,t} &= \frac{C_{i,t} - C_{i,t-1}}{Y_{i,0}} \\
\Delta \tilde{y}_{i,t} &= \frac{Y_{i,t} - Y_{i,t-1}}{Y_{i,0}} \\
\Delta \tilde{w}^h_{i,t} &= \frac{W^h_{i,t} - W^h_{i,t-1}}{Y_{i,0}} \\
\Delta \tilde{w}^f_{i,t} &= \frac{(W^f_{i,t} - W^f_{i,t-1})}{Y_{i,0}},
\end{aligned}$$

where $Y_{i,0}$ is the state after-tax labor income at 2000h1, then the following regression

$$\Delta \tilde{c}_{i,t} = \alpha_t + \beta_1 \Delta \tilde{y}_{i,t} + \beta_2 \Delta \tilde{w}^f_{i,t} + \beta_3 \Delta \tilde{w}^h_{i,t} + \Delta \tilde{\varepsilon}_t, \tag{4.2}$$

will potentially produce direct measures of the MPC out of the changes in housing wealth and stock wealth.

As with Equation 4.1, Equation 4.2 is subject to serious endogeneity problems, and thus is considered as simply another data description. Table 4.2 indicates that under this model specification, income change is still the most correlated variable with respect to consumption.

In order to resolve the endogeneity and simultaneity problem that Equation 4.2 is subject to, we briefly revisit classic consumption theory. The relationship between consumption and wealth/income can be described by the Life-Cycle/Permanent Income Hypothesis.

[11] One benefit of such estimations is that they produce certain results comparable to those in the current literature. For the sake of comparison, the results of similar estimations are included in the appendix of this paper.

Specifically, a consumer wants to

$$\text{MAX } E_t[\sum_{s=t}^{\infty} \beta_{s-t} u(c_s)]$$

subject to the budget constraint, where β is the time preference, and $u(c_t)$ is the utility function. If the utility function takes a quadratic form as assumed in Hall (1978), it can be easily shown that, under certain conditions, consumption will follow a random walk, i.e.,

$$\Delta c_{t+1} = \epsilon_{t+1},$$

$$E_t[\epsilon_{t+n}] = 0 \ \forall n > 0$$

Thus, the theory implies that consumption responds to unexpected shocks only. In other words, information known to consumers at the time when consumption choices are made cannot have any predictive power for consumption changes in any future periods.

The random walk proposition, therefore, can help us alleviate the endogeneity and simultaneity problem, as it suggests that current consumption growth would not react to any lagged wealth growth. Nevertheless, time aggregation and measurement error could cause current consumption changes to correlate with once lagged income and wealth changes, even if the PIH holds true. Aggregation also matters when the PIH holds in continuous time, and the measures of consumption are based on time averages. Under this situation, changes in time-averaged consumption will have nonzero first order serial correlations; this will lead to nonzero correlations between changes in consumption and once-lagged variables. It is also easy to prove that measurement errors in the consumption level could cause measured consumption changes that correlate with once-lagged explanatory variables.[12] Given the above considerations, the following equation is employed to address the question of wealth

[12]Let us assume that $c_t = c_{t-1} + \varepsilon_t$ and $c_t = c_t^* + \upsilon_t$, where c_t is real consumption, c_t^* is the measured consumption, and υ_t is the measurement error. Although real consumption growth follows a random walk, the measured consumption growth, $\Delta c_t^* = \varepsilon_t - (\upsilon_t - \upsilon_{t-1})$, is correlated with the once-lagged information.

effects[13]:

$$\Delta\tilde{c}_{i,t} = \alpha_t + \beta_1 \Delta\tilde{y}_{i,t-2} + \beta_2 \Delta\tilde{w}^f_{i,t-2} + \beta_3 \Delta\tilde{w}^h_{i,t-2} + \Delta\tilde{\varepsilon}_t. \tag{4.3}$$

Equation 4.3 employs twice-lagged independent variables, and thus reports MPCs out of changes in housing wealth and stock wealth that occurred two periods prior.

There are, however, two minor modifications that need to be made. First, what $C_{i,t}$ captures here is not the real personal consumption for state i, but the state's taxable retail sales. Thus, using $C_{i,t}$, the estimation of Equation 4.3 actually yields the effect of changes in wealth on taxable retail sales. To gauge the approximate change in real consumption, it is assumed that initial state consumption can be determined by $C^*_{i,0} = Y_{i,0} * \frac{C^*_0}{Y_0}$, where C^*_0 and Y_0 are aggregate personal consumption expenditure and after-tax labor income, respectively. In addition, we assume that the ratio of retail sales to real consumption holds constant over time, i.e., $\frac{C_{i,t}}{C^*_{i,t}} = \frac{C_{i,0}}{C^*_{i,0}}$. Therefore, changes in state consumptions can be measured roughly by

$$\begin{aligned}
(C^*_{i,t} - C^*_{i,t-1}) &= (C_{i,t} - C_{i,t-1}) * (\frac{C^*_{i,0}}{C_{i,0}}) \\
&= (C_{i,t} - C_{i,t-1}) * (\frac{C^*_0}{Y_0}\frac{Y_{i,0}}{C_{i,0}}).
\end{aligned}$$

The same problem arises when measuring stock wealth. Thus, it is assumed for all states and time periods that $\frac{W^f_{i,t}}{W^{f*}_{i,t}} = \frac{W^f_{IXI,0}}{W^f_{FFA,0}}$, where $W^{f*}_{i,t}$ denotes the real state stock wealth at time t.

[13]IV regression is another commonly used method to solve endogeneity problems. However, variables lagged by two years have weak explanatory power, especially for income and stock wealth growth. Thus, it would lead us to another econometric issue – weak instruments.

Therefore, if we redefine

$$\Delta\tilde{c}_{i,t} = \frac{C_{i,t} - C_{i,t-1}}{Y_{i,0}} * (\frac{C_0^*}{Y_0}\frac{Y_{i,0}}{C_{i,0}}) \tag{4.4}$$

$$\Delta\tilde{w}_{i,t}^f = \frac{(W_{i,t}^f - W_{i,t-1}^f)}{Y_{i,0}} * \frac{W_{FFA,0}^f}{W_{IXI,0}^f}, \tag{4.5}$$

the regression of Equation 4.3 ends up reporting approximate estimates of the MPC out of changes in housing wealth and stock wealth.

Table 4.3 summarizes the results of our estimations using Equation 4.3. It indicates that all three datasets report similar results, with the exception that none of the estimations from "Best Data" is statistically significant. It is, however, well expected, given the small sample size of "Best Data."

Table 4.3 shows that the coefficients of income changes are all positive and large. Furthermore, they are statistically significant by using both "Combined Data" and "Good Data". It therefore implies that income changes have a fairly big impact on consumption, despite the two-year lag. This, however, contradicts the random walk theory as predicted by the Permanent Income Hypothesis.

The wealth effect caused by changes in financial wealth, on the other hand, is found to be both significant and negligible. This finding is consistent with Dynan and Maki (2001), who found that the impact of stock wealth on consumption very quickly becomes apparent, and any lagged changes in stock wealth beyond 9 months does not have any significant effect on consumption.

However, we observe highly significant and large coefficients for housing wealth in two out of the three datasets. Additionally, all three datasets indicate that an MPC out of housing wealth changes occurs two years prior around the neighborhood of 6 cents. The main reasons why the response to housing wealth shocks may be slower than the response to financial wealth shocks are: Unlike stock prices that can be easily tracked daily online or in newspapers, house prices cannot be accurately and regularly observed. Actually, homeowners might be less aware of short-run changes in house prices and it might take a homeowner a while to realize that his/her house price has changed. Additionally, the cost of realizing capital gains on housing wealth is lumpy. As a result, the response to housing wealth growth is not likely to be spontaneous.

What is more interesting is that the difference between the housing wealth effect and the stock wealth effect is found to be statistically significant for "Good Data," and on the verge of being significant for "Combined Data." Therefore, in the presence of the consistently larger point estimates for the housing wealth effect, when implementing policies and making macroeconomic forecasts, monetary policymakers should be alert to the different impacts on consumption generated by movements in the housing and stock markets respectively.

4.5.2 A habit formation test

The above estimations only report the relatively immediate impact of wealth changes on consumption. In place of a cointegration analysis, this paper applies a method proposed by Carroll, Otsuka, and Slacalek (2006) for deriving "long term" wealth effects. The basic idea is, if there is evidence of habit formation, consumption growth will be serially correlated. Thus, any impact that wealth changes have on consumption could be delivered over the very long run through a serial correlation of consumption growth. The long run wealth effect then can be derived by dividing the short run wealth effect by one minus the habit formation coefficient. Following the relevant literature, the following equation is employed as a habit formation test:

$$\Delta\tilde{c}_{i,t} = \alpha_t + \lambda E_{t-2}\Delta\tilde{c}_{i,t-1}. \tag{4.6}$$

Table 4.4 reports the estimations using Equation 4.6. Using currently available state-level instruments, the results provide no evidence of habit formation.[14] This could be because of the short time horizon of the data. Consequently, the estimation for habit formation

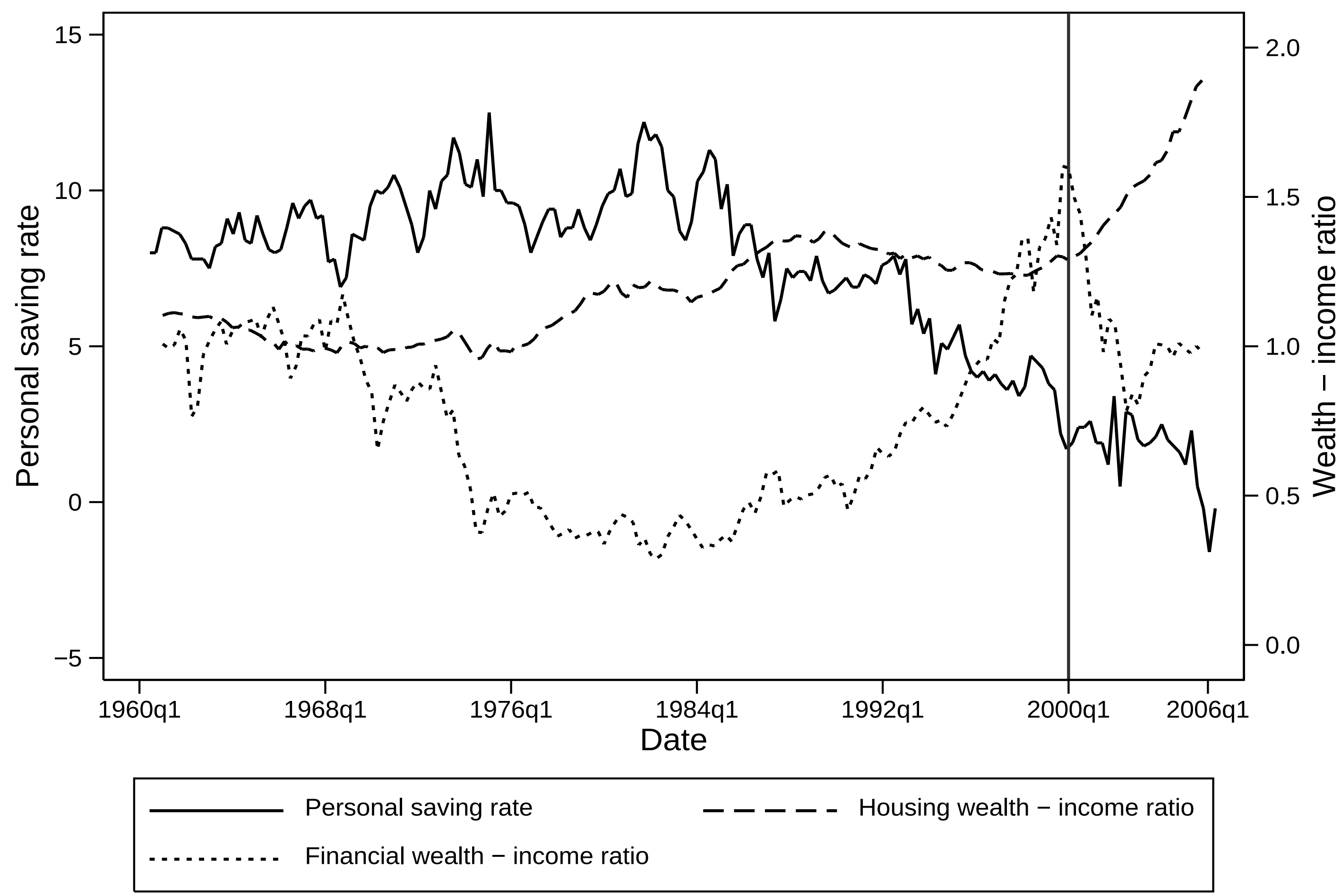
[14]Many other IV sets were tested but also failed to demonstrate a positive and significant habit formation coefficient with reasonable first stage results.

remains as an interesting topic for future research.

4.6 Conclusion

This paper describes a new panel dataset of financial wealth for U.S. states, constructed from anonymous proprietary account-level records of geographic wealth holdings. The new dataset is more comprehensive and representative than existing alternative measures. The paper also constructs significantly improved state-level consumption data, and then combines these datasets to provide new estimates of the effects of changes in stock wealth and housing wealth on consumption. Consistent and strong evidence is found for large but sluggish housing wealth effects. Based on the results from our new approach, two out of the three datasets indicate that the MPC out of a one dollar change in two-year lagged housing wealth is about 6 cents. In addition, the twice-lagged income change is also found to have large impact on current consumption. Both findings lead to the rejection of the random walk theory. Furthermore, a statistically insignificant and economically small stock wealth effect is found for almost all specifications. Additionally, there is evidence that the housing wealth effect is significantly larger than the stock wealth effect. These results could, nonetheless, help explain the strength of consumption following the stock market bubble burst at the end of the 1990s. With respect to monetary policies then, these results suggest that it is necessary to take into consideration the potentially substantial difference between consumers' respective reactions to fluctuations in the housing markets and stock markets.

Figure 4.1: The saving rate versus the net worth - income ratio



Source: BEA and FFA

Figure 4.2: Mutual funds versus total stock wealth

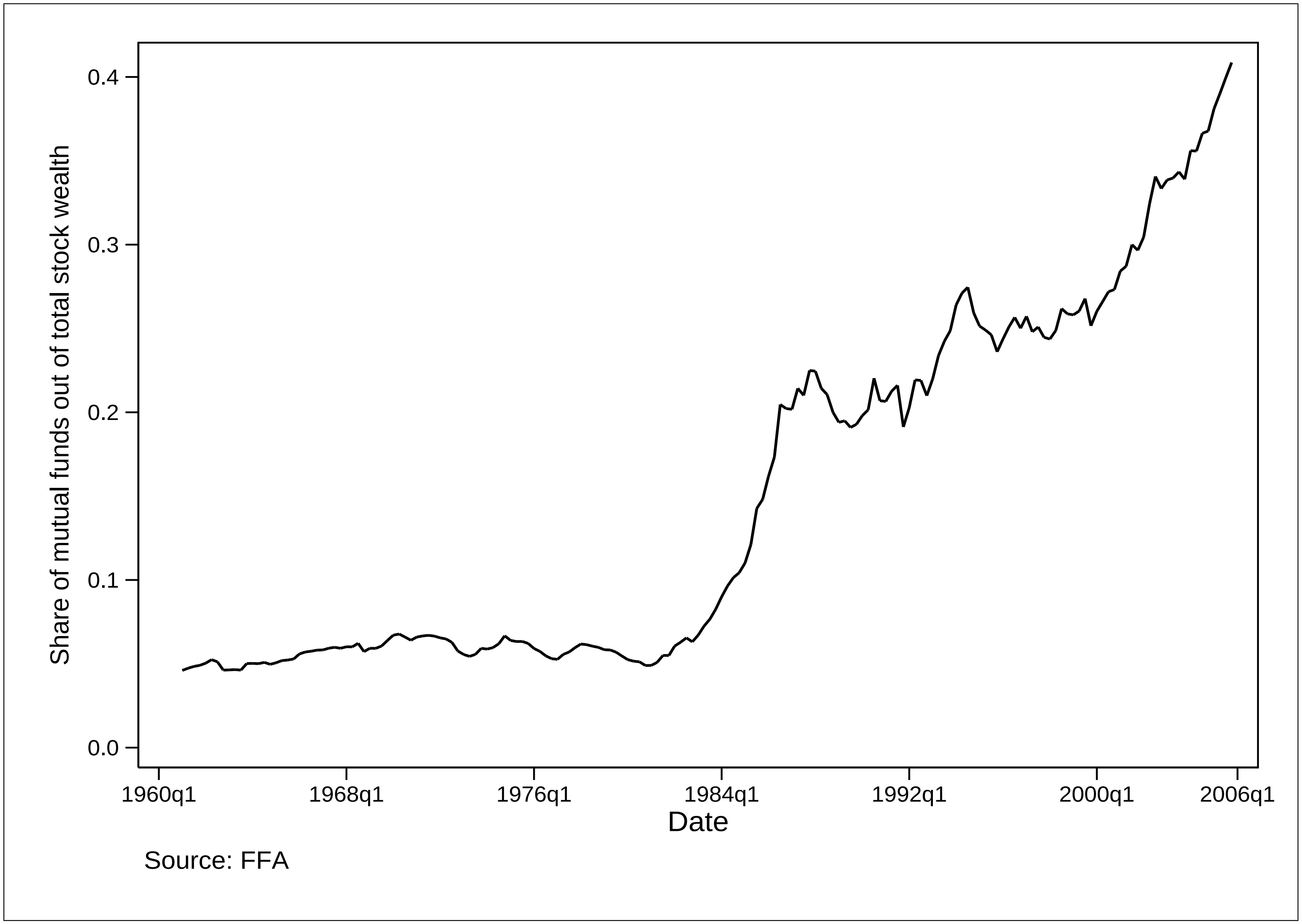
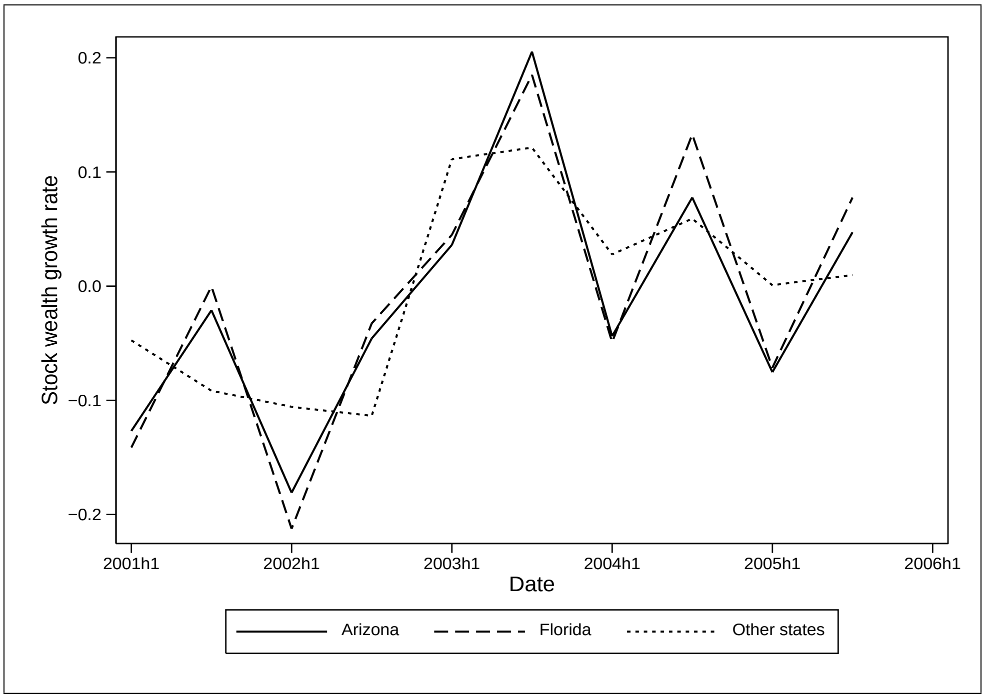
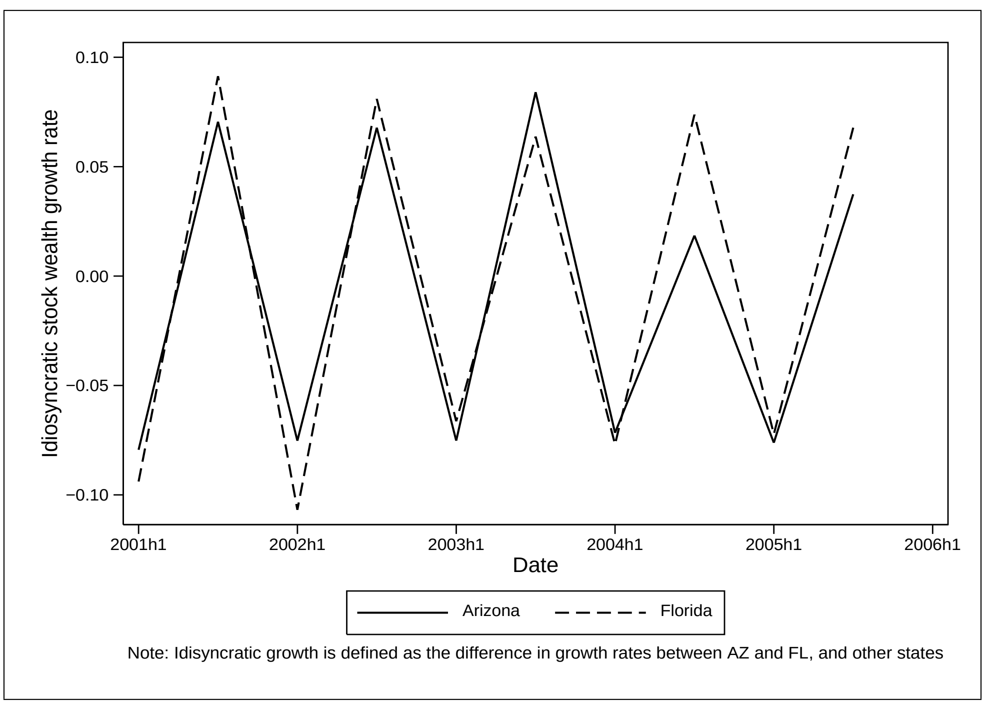


Figure 4.3: Snow bird effect



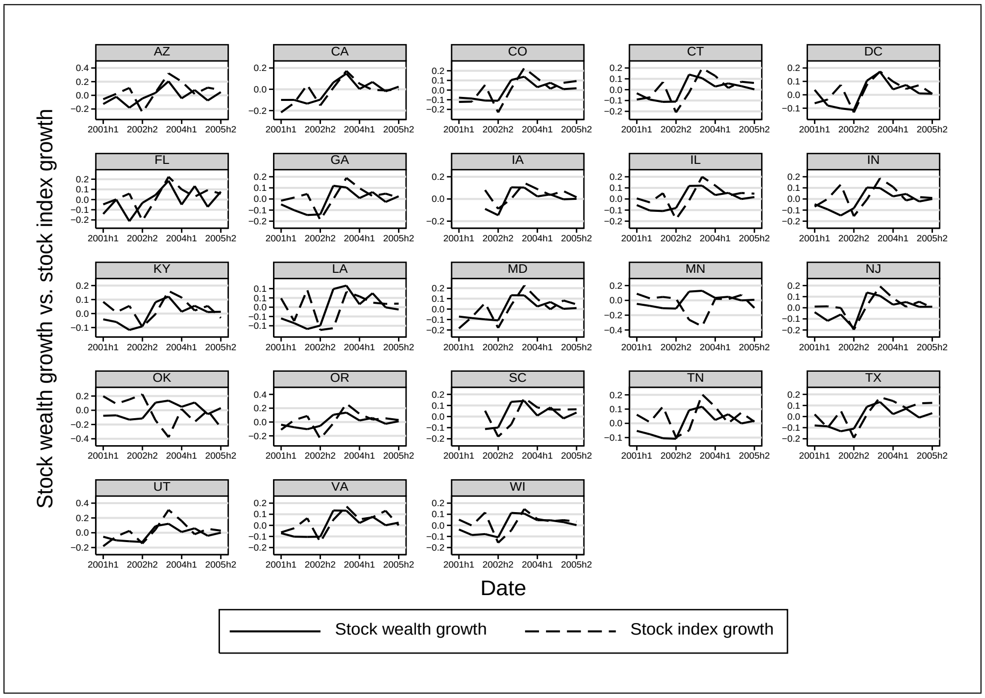
(a) Florida and Arizona versus the average of other states



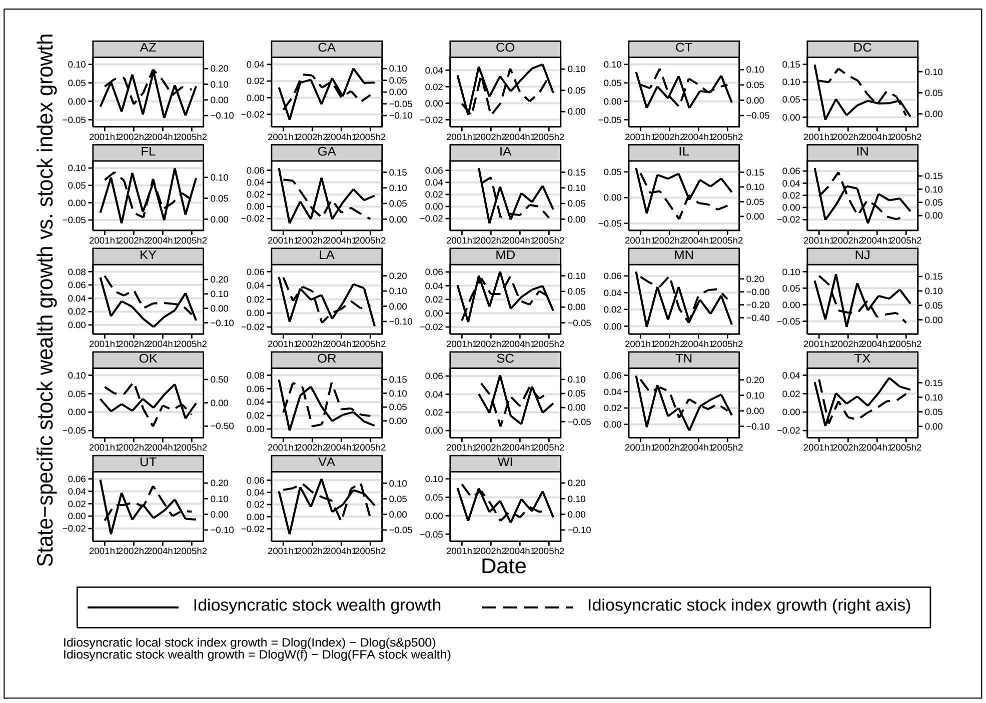
(b) Florida and Arizona versus other states

Note: The sharp seasonal fluctuations in wealth in Florida and Arizona likely reflect a ”snow bird” effect, as wealthy retirees move in and out of these states on a seasonal basis.

Figure 4.4: The local stock index versus state stock wealth



(a) State financial wealth growth versus local stock index growth



(b) Idiosyncratic state financial wealth growth versus idiosyncratic local stock index growth

Table 4.1: Data description: $\Delta c_{i,t} = \alpha_t + \beta_1 \Delta y_{i,t} + \beta_2 \Delta w^f_{i,t} + \beta_3 \Delta w^h_{i,t}$

Best Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Delta y_{i,t}$	0.766***			0.68***	0.793***		0.704***
	(0.204)			(0.202)	(0.191)		(0.184)
$\Delta w^f_{i,t}$		0.43**		0.352**		0.449***	0.369**
		(0.176)		(0.176)		(0.168)	(0.163)
$\Delta w^h_{i,t}$			0.125*		0.135**	0.134**	0.141**
			(0.064)		(0.059)	(0.063)	(0.059)
Obs.	48	48	48	48	48	48	48
$\bar{R}^2$	0.72	0.701	0.687	0.739	0.743	0.722	0.765
Partial $\bar{R}^2$	0.154	0.095	0.051	0.212	0.222	0.16	0.291

Combined Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Delta y_{i,t}$	1.945***			1.844**	1.936***		1.851**
	(0.698)			(0.721)	(0.736)		(0.752)
$\Delta w^f_{i,t}$		0.392**		0.293		0.376**	0.294
		(0.19)		(0.249)		(0.189)	(0.248)
$\Delta w^h_{i,t}$			0.107		0.011	0.077	-.008
			(0.078)		(0.087)	(0.074)	(0.081)
Obs.	180	180	180	180	180	180	180
$\bar{R}^2$	0.21	0.126	0.102	0.222	0.206	0.124	0.217
Partial $\bar{R}^2$	0.121	0.027	0.0008	0.135	0.116	0.025	0.13

Good Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Delta y_{i,t}$	1.945***			1.844**	1.936***		1.851**
	(0.698)			(0.721)	(0.736)		(0.752)
$\Delta w^f_{i,t}$		0.392**		0.293		0.376**	0.294
		(0.19)		(0.249)		(0.189)	(0.248)
$\Delta w^h_{i,t}$			0.107		0.011	0.077	-.008
			(0.078)		(0.087)	(0.074)	(0.081)
Obs.	180	180	180	180	180	180	180
$\bar{R}^2$	0.21	0.126	0.102	0.222	0.206	0.124	0.217
Partial $\bar{R}^2$	0.121	0.027	0.0008	0.135	0.116	0.025	0.13

a. Partial $\bar{R}^2$ refers to the proportion of variance explained by all variables other than the year dummies.
b. Standard errors in parenthesis. {*, **, ***} = significant at the {10%, 5%, 1%} level.

Table 4.2: $\Delta\tilde{c}_{i,t} = \alpha_t + \beta_1 \Delta\tilde{y}_{i,t} + \beta_2 \Delta\tilde{w}^f_{i,t} + \beta_3 \Delta\tilde{w}^h_{i,t}$

	Best Data	Combined Data	Good Data
$\Delta y_{i,t}$	0.76***	2.509**	1.519***
	(0.25)	(1.095)	(0.537)
$\Delta w^f_{i,t}$	0.073**	0.023	0.042
	(0.029)	(0.059)	(0.043)
$\Delta w^h_{i,t}$	0.016	0.006	0.012
	(0.01)	(0.013)	(0.01)
$\beta_2 = \beta_3$	3.555	0.088	0.473
	(Rejected)	(Accepted)	(Accepted)
OBS	48	180	180
$\bar{R}^2$	0.767	0.201	0.251
Partial $\bar{R}^2$	0.309	0.127	0.111

Table 4.3: $\Delta\tilde{c}_{i,t} = \alpha_t + \beta_1 \Delta\tilde{y}_{i,t-2} + \beta_2 \Delta\tilde{w}^f_{i,t-2} + \beta_3 \Delta\tilde{w}^h_{i,t-2}$

	Best Data	Combined Data[a]	Good Data
$\Delta y_{i,t-2}$	0.556	1.083**	0.891**
	(0.423)	(0.423)	(0.395)
$\Delta w^f_{i,t-2}$	-.005	-.015	-.021
	(0.04)	(0.035)	(0.029)
$\Delta w^h_{i,t-2}$	0.067	0.058**	0.061***
	(0.048)	(0.025)	(0.022)
$\beta_2 = \beta_3$	1.666	2.688	4.956**
	(Accepted)	(Accepted)	(Rejected)
OBS	24	90	90
$\bar{R}^2$	0.244	0.03	0.061
Partial $\bar{R}^2$	0.037	0.039	0.072

[a] The regression using only the computed taxable retail sales shows a significant and even larger housing wealth effect, and a very small and insignificant financial wealth effect. The results are available from the author upon request.

Table 4.4: Habit formation: $\Delta\tilde{c}_{i,t} = \alpha_t + \lambda E_{t-2}\Delta\tilde{c}_{i,t-1} + \varepsilon_t$

	Best Data	Combined Data	Good Data
$E_{t-2}\Delta\tilde{c}_{i,t-1}$[a]	0.642	-.004	0.074
	(0.4)	(0.301)	(0.314)
obs	24	90	90
$\bar{R}^2$	0.028	-.018	-.017
First Stage:			
Partial R^2	0.33	0.156	0.139
$P-val$	0.069	0.0005	0.017

[a] IV: $\Delta\tilde{c}_{i,t-2}$, $\Delta\tilde{y}_{i,t-2}$, $\Delta\tilde{w}^f_{i,t-2}$, $\Delta\tilde{w}^h_{i,t-2}$.

APPENDIX: Results using the elasticity method

Many papers in the literature have estimated wealth effects by adopting the elasticity method. Consequently, we then investigate the respective housing wealth and stock wealth effects by estimating the following equation, as with most related studies:

$$\Delta c_{i,t} = \alpha_t + \beta_1 \Delta y_{i,t-2} + \beta_2 \Delta w^f_{i,t-2} + \beta_3 \Delta w^h_{i,t-2} + \varepsilon_{i,t}. \tag{4.7}$$

Table 4.5 reports the regression results from Equation 4.7 for all three sets of consumption data. The findings are roughly consistent across the three datasets.

The most outstanding and robust finding is the large coefficient for lagged housing wealth. The stock wealth effects reported in Table 4.5 are all statistically insignificant. Furthermore, in 2 of the 3 estimations, the size of the stock wealth effect is economically small. The hypotheses of equal housing wealth and stock wealth coefficients are, however, accepted in 2 out of 3 estimations.

Table 4.5: Results for the elasticity method

	Best Data	Combined Data	Good Data
$\Delta y_{i,t-2}$	0.338 (0.321)	0.609* (0.312)	0.405 (0.286)
$\Delta w^f_{i,t-2}$	0.234 (0.269)	-.022 (0.098)	-.072 (0.084)
$\Delta w^h_{i,t-2}$	0.411** (0.198)	0.266** (0.115)	0.278*** (0.099)
Test of β_2=β_3	0.243 (Accepted)	2.614 (Accepted)	5.61** (Rejected)
obs	24 90	90	
$\bar{R}^2$	0.37	0.015	0.042
Partial $\bar{R}^2$	0.177	0.024	0.052

Bibliography

Antoniewicz, R. L. (2000): "A Comparison of the Household Sector from the Flow of Funds Accounts and the Survey of," Discussion paper, Federal Reserve Board.

Antonio, C., and R. E. Hall (1996): "Productivity and the Density of Economic Activity," *The American Economic Review*, 86(1), 54–70.

Aron, J., J. V. Duca, J. Muellbauer, K. Murata, and A. Murphy (2010): "Credit, Housing Collateral and Consumption: Evidence from the UK, Japan and the US," Working Paper 1002, Federal Reserve Bank of Dallas.

Aron, J., and J. Muellbauer (2006): "Housing Wealth, Credit Conditions and Consumption," Centre for the Study of African Economies Series WPS 2006-08, University of Oxford.

Arrow, K. J. (1962): "The Economic Implications of Learning by Doing," *Review of Economic Studies*, 29(3), 155–173.

Asdrubali, P., B. Sorensen, and O. Yosha (1996): "Channels of interstate risksharing: United States 1963-1990," *Quarterly Journal of Economics*, 111, 1081–1110.

Bayoumi, T., and H. Edison (2003): "Is Wealth Increasingly Driving Consumption," DNB Staff Reports (discontinued) 101, Netherlands Central Bank.

Bertaut, C. (2002): "Equity Prices, Household Wealth, and Consumption Growth in Foreign Industrial Countries: Wealth Effects in the 1990s," International Finance Discussion Papers 724, Board of Governors of the Federal Reserve System.

Bostic, R., S. Gabriel, and G. Painter (2005): "Housing Wealth, Financial Wealth, and Consumption: New Evidence from Micro Data," manuscript, University of Southern California.

Byrne, J. P., and E. P. Davis (2003): "Disaggregate Wealth and Aggregate Consumption: an Investigation of Empirical Relationships for the G7," *Oxford Bulletin of Economics and Statistics*, 65(2), 197–220.

Campbell, J. Y., and J. F. Cocco (2006): "How Do House Prices Affect Consumption? Evidence From Micro Data," *Journal of Monetary Economics*, forthcoming.

Campbell, J. Y., and G. N. Mankiw (1989): "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER macroeconomics annual*, ed. by O. J. Blanchard, and S. Fischer, pp. 185–216. MIT Press, Cambridge, MA.

Cardarelli, R., D. Igan, and A. Rebucci (2008): "The Changing Housing Cycle and the Implications for Monetary Policy," in *World Economic Outlook*, chap. 3. Washington: International Monetary Fund.

Carroll, C. D., M. Otsuka, and J. Slacalek (2006): "How Large Is the Housing Wealth Effect? A New Approach," Working paper no. 535, Johns Hopkins University, www.econ.jhu.edu/pdf/papers/wp535carroll.pdf.

Case, K. E., J. M. Quigley, and R. J. Shiller (2005): "Comparing Wealth Effects: The Stock Market versus The Housing Market," *Advances in Macroeconomics*, 5(1), 1–34.

Ciccone, A. (2002): "Agglomeration effects in Europe," *European Economic Review*, 46(2), 213–227.

Curtin, R. T., F. T. Juster, and J. N. Morgan (1989): "Survey Estimates of Wealth: An Assessment of Quality," *The Measurement of Saving, Investment and Wealth*, 52, 473–548.

Davis, M., and M. G. Palumbo (2001): "Does Stock Market Wealth Matter for Consumption?," Finance and Economics Discussion Series 2001-23. Washington: Board of Governors of the Federal Reserve System.

Del Negro, M. (1998): "Aggregate Risk Sharing Across US States and Across European Countries," Ph.D. thesis, Yale University.

Disney, R., A. Henley, and D. Jevons (2003): "House Price Shocks, Negative Equity and Household Consumption in the UK in the 1990s," memo, University of Nottingham.

Dvornak, N., and M. Kohler (2003): "Housing Wealth Stock Market Wealth and Consumption: A Panel Analysis for Australia," RBA Research Discussion Papers rdp2003-07, Reserve Bank of Australia.

Dynan, K. E., and D. M. Maki (2001): "A Primer on the Economics and Time Series Econometrics of Wealth Effects," Finance and Economics Discussion Series 2001-9. Washington: Board of Governors of the Federal Reserve System.

Edelstein, R. H., and S. K. Lum (2004): "House prices, wealth effects, and the Singapore macroeconomy," *Journal of Housing Economics*, 13, 342–367.

Garrett, T. A., R. Hernàndez-Murillo, and M. T. Owyang (2004): "Does Consumer Sentiment Predict Regional Consumption?," *Federal Reserve Bank of St. Louis Review*, 87(2), 123–35.

Girouard, N., and S. Blöndal (2001): "House Prices and Economic Activity," OECD Economics Department Working Papers 279, OECD Economics Department.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992): "Growth in Cities," *Journal of Political Economy*, 100(6), 1126–52.

Green, R. K. (2002): "Stock prices and house prices in California: new evidence of a wealth effect?," *Regional Science and Urban Economics*, 32, 775–783.

Hall, R. E. (1978): "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86(6), 971–87.

Hess, G. D., and K. Shin (1998): "Intranational business cycles in the United States," *Journal of International Economics*, 44, 289–313.

Jacobs, J. (1969): *The Economy of Cities*. New York: Vintage.

KENNICKELL, A. B. (1999): "Revisions to the SCF Weighting Methodology: Accounting for Race/Ethnicity and Homeownership," Discussion paper, Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/papers/weight.revision.pdf.

——— (2003): "A Rolling Tide: Changes in the Distribution of Wealth in the U.S., 1989-2001," Discussion paper, Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/papers/concentration.2004.5.pdf.

——— (2006a): "Currents and Undercurrents: Changes in the Distribution of Wealth, 1989-2004," Discussion paper, Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/papers/concentration.2001.10.pdf.

——— (2006b): "How Do We Know If We Aren't Looking? An Investigation Of Data Quality in the 2004 SCF," Discussion paper, Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/papers/asa2006.3.pdf.

——— (2006c): "List Sample Design for 1998 Survey of Consumer Finances," Discussion paper, Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/papers/listsample.pdf.

LETTAU, M., AND S. LUDVIGSON (2001): "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, LVI(3), 815–849.

——— (2004): "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption," *American Economic Review*, 94(1), 279–299.

LEVIN, L. (1998): "Are assets fungible? Testing the behavioral theory of life-cycle savings," *Journal of Economic Organization and Behavior,*, 36, 59–83.

LOPEZ, M. A. G., AND I. M. OLIVERA (2005): "The Spatial Effect of Intra-Metropolitan Agglomeration Economies," Working Papers 0513, Department of Applied Economics at Universitat Autonoma of Barcelona.

LUDWIG, A., AND T. SLØK (2002): "The Impact of Changes in Stock Prices and House Prices on Consumption in OECD Countries," Working Paper 02/1 International Monetary Fund.

Luengo-Prado, M. J., and B. Sorensen (2006): "What Can Explain Excess Smoothness and Sensitivity of State-Level Consumption?," *Forthcoming in the Review of Economics and Statistics*.

Mian, A., and A. Sufi (2009): "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics*, 124(4).

Moore, K., and B. Johnson (2005): "Consider the Source: Differences in Estimates of Income and Wealth from Survey and Tax Data," Discussion paper, Federal Reserve Board, Internal Revenue Service, http://www.federalreserve.gov/pubs/oss/oss2/papers/johnsmoore.pdf.

Porter, M. E. (1990): *The Competitive Advantage of Nations*. New York: Free Press.

Ravina, E. (2005): "Habit Persistence and Keeping Up with the Joneses: Evidence from Micro Data," Working Paper.

Romer, P. M. (1986): "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 94(5), 1002–37.

Shiller, R. J. (2004): "Household Reaction to Chagnes in Housing Wealth," Discussion paper, Cowles Foundation for Research in Economics, Yale University, Discussion Paper NO. 1459.

Thaler, R. H. (1990): "Anomalies: Saving, Fungibility, and Mental Accounts," *Journal of Economic Perspectives*, 4(1), 193–205.

Wagner, A. (1891): "Principles of Economics," *History of Economic Thought Articles*, 5, 319–338.

Wilcox, D. W. (1992): "The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work," *The American Economic Review*, 82(4), 922–941.

Zhu, X. (2007): "Growing Wealth, Inequality, and Housing in the United States," Working Paper W07-1, Joint Center for Housing Studies, Harvard University.

XIA ZHOU

Department of Economics
Johns Hopkins University
3400 North Charles Street
Baltimore, MD 21218

7252 Swansong Way
Bethesda, MD 20817
Phone: 443-858-2697
E-mail: xzhou@jhu.edu

PERSONAL INFORMATION

Date and Location of Birth: January 18th, 1977, Hubei, China
Citizenship: P.R.China

EDUCATION

Ph.D. in Economics, Johns Hopkins University, 2010

M.A. in Economics, Johns Hopkins University, 2003

M.A. in Economics, Hong Kong University of Science and Technology, 2001

B.A. in International Economics and Trade, Nankai University, 1998

FIELDS OF INTEREST

Consumption, Macroeconomics, Applied Macro

RESEARCH PAPERS

"Measuring Wealth Effects Using U.S. States Data"

"Study of the Efficacy of Interest Rate Defense against Currency Speculations: The Case of Hong Kong"

HONORS AND FELLOWSHIPS

Fellowship, Department of Economics, Johns Hopkins University, 2002 – 2006

Fellowship, Department of Economics, Hong Kong University of Science and Technology, 1999 – 2001

Excellent Graduation Thesis Award, Department of International Economics and Trade, Nankai University, 1998

Scholarship, Department of International Economics and Trade, Nankai University, 1994 – 1996